

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号

特許第7121440号

(P7121440)

(45)発行日 令和4年8月18日(2022.8.18)

(24)登録日 令和4年8月9日(2022.8.9)

(51)国際特許分類

F I

G 1 6 B 40/10 (2019.01)

G 1 6 B 40/10

請求項の数 47 (全74頁)

(21)出願番号	特願2022-521759(P2022-521759)	(73)特許権者	520496501
(86)(22)出願日	令和3年12月16日(2021.12.16)		株式会社 s e e D N A
(86)国際出願番号	PCT/JP2021/046513		東京都足立区竹の塚3-10-1 竹の
審査請求日	令和4年4月11日(2022.4.11)		塚ビル2階
(31)優先権主張番号	特願2020-208554(P2020-208554)	(74)代理人	100137338
(32)優先日	令和2年12月16日(2020.12.16)		弁理士 辻田 朋子
(33)優先権主張国・地域又は機関	日本国(JP)	(74)代理人	100196313
早期審査対象出願			弁理士 村松 大輔
		(72)発明者	金 起範
			東京都足立区竹の塚3-10-1 竹の
			塚ビル2階 株式会社 s e e D N A 内社
			横浜研究所内
		(72)発明者	杉元 崇紀
			東京都足立区竹の塚3-10-1 竹の
			塚ビル2階 株式会社 s e e D N A 内社
			最終頁に続く

(54)【発明の名称】 多型座位の信号の信頼性値の算出方法

(57)【特許請求の範囲】

【請求項1】

以下の工程A-1、工程A-2、工程A-3-1及び工程A-4-1を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

【工程A-1】

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

【工程A-2】

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の(A1)及び(A2)を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

(A1) 前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度

(A2) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率

【工程A-3-1】

前記工程A-2で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数

に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

【工程 A-4-1】

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【請求項 2】

前記合成変数が、少なくとも前記 (A 1) 及び前記 (A 2) を含む数値群について主成分分析を行うことで生成し得る合成変数であることを特徴とする、請求項 1 に記載の方法。

【請求項 3】

前記工程 A-3-1 及び工程 A-4-1 においてモデル関数の作成のために用いる前記合成変数が、工程 A-2 において生成した 1 以上の合成変数のうち、最も寄与率が高い合成変数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】

前記工程 A-2 が、少なくとも前記 (A 1) 及び前記 (A 2) を含む数値群について主成分分析を行い、1 つ以上の主成分を合成変数として生成する工程であることを特徴とする、請求項 1 ～ 3 の何れか一項に記載の方法。

【請求項 5】

前記工程 A-2 が、前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A 1) 及び前記 (A 2) を含み、さらに以下の (A 3) ～ (A 5) から選ばれる 1 又は 2 以上を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程であることを特徴とする、請求項 1 ～ 4 の何れか一項に記載の方法。

(A 3) 前記主要核酸に由来する、特定の多型座位の一のアレルの存在を示す主要成分信号強度。

(A 4) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記主要成分信号強度の割合である、主要成分混入率。

(A 5) 前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

【請求項 6】

前記工程 A-2 が、前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A 1) 及び (A 2) を含み、さらに以下の (A 3) ～ (A 5) を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程であることを特徴とする、請求項 1 ～ 5 の何れか一項に記載に記載の方法。

(A 3) 前記主要核酸に由来する、特定の多型座位の一のアレルの存在を示す主要成分信号強度。

(A 4) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記主要成分信号強度の割合である、主要成分混入率。

(A 5) 前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

【請求項 7】

前記回帰分析が最小二乗法であることを特徴とする、請求項 1 ～ 6 の何れか一項に記載の方法。

【請求項 8】

前記合成変数を表す 1 次斉次多項式において、副次成分信号強度又は副次成分混入率が最大に重みづけされることを特徴とする、請求項 1 ～ 7 の何れか一項に記載の方法。

【請求項 9】

前記工程 A-2 において線形結合する数値群に含まれる数値は、標準化された数値であることを特徴とする、請求項 1～8 の何れか一項に記載の方法。

【請求項 10】

前記工程 A-2 において、2 つ以上の合成変数を生成し、

前記工程 A-3-1 において、前記 2 つ以上の合成変数のそれぞれについて、信頼性値の付与を行い、

前記工程 A-4-1 において、前記 2 つ以上の合成変数のそれぞれを説明変数とする互いに独立した 2 以上のモデル関数を作成し、

さらに、前記 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数を作成する工程を備えることを特徴とする、請求項 1～9 の何れか一項に記載の方法。

10

【請求項 11】

以下の工程 A-1、工程 A-3-2 及び工程 A-4-2 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

【工程 A-1】

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

【工程 A-3-2】

前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度を複数に区分し、各区分に含まれる前記副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記副次成分信号強度に対応する確率として付与する工程。

20

【工程 A-4-2】

前記各区分に含まれる前記副次成分信号強度と、前記各区分に含まれる前記副次成分信号強度に対応する確率について回帰分析を行い、前記副次成分信号強度を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【請求項 12】

以下の工程 A-1、工程 A-3-3 及び工程 A-4-3 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

30

【工程 A-1】

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

【工程 A-3-3】

前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合である副次成分混入率を複数に区分し、各区分に含まれる前記副次成分混入率に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記副次成分混入率に対応する確率として付与する工程。

40

【工程 A-4-3】

前記各区分に含まれる前記副次成分混入率と、前記各区分に含まれる前記副次成分混入率に対応する確率について回帰分析を行い、前記副次成分混入率を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【請求項 13】

前記モデル関数が、シグモイド関数であることを特徴とする、請求項 1～12 の何れか一項に記載の方法。

50

【請求項 14】

前記モデル関数が、2 個の媒介変数を有するシグモイド関数であることを特徴とする、請求項 1 ～ 13 の何れか一項に記載の方法。

【請求項 15】

請求項 1 ～ 10 の何れか一項に記載の方法で作成したモデル関数と、
請求項 11 に記載の方法で作成したモデル関数と、
請求項 12 に記載の方法で作成したモデル関数と、
からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【請求項 16】

請求項 1 ～ 10 の何れか一項に記載の方法で作成したモデル関数と、
請求項 11 に記載の方法で作成したモデル関数、及び／又は、請求項 12 に記載の方法で作成したモデル関数と、
を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【請求項 17】

請求項 1 ～ 10 の何れか一項に記載の方法で作成したモデル関数と、
請求項 11 に記載の方法で作成したモデル関数と、
請求項 12 に記載の方法で作成したモデル関数と、
を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【請求項 18】

前記多型座位が、一塩基多型を含む座位であることを特徴とする、請求項 1 ～ 17 の何れか一項に記載の方法。

【請求項 19】

前記データセットが、塩基配列解析、デジタル PCR、マイクロアレイ、リアルタイム PCR 又は質量分析により取得されたデータであることを特徴とする、請求項 1 ～ 18 の何れか一項に記載の方法。

【請求項 20】

前記データセットが塩基配列解析により取得されたデータであり、
前記副次成分信号強度が、配列タグのカウント数、リード数、イオン濃度又は電気信号であることを特徴とする、請求項 1 ～ 18 の何れか一項に記載の方法。

【請求項 21】

前記データセットがデジタル PCR により取得されたデータであり、
前記副次成分信号強度が、蛍光が観察されたウェル数であることを特徴とする、請求項 1 ～ 18 の何れか一項に記載の方法。

【請求項 22】

前記データセットがマイクロアレイにより取得されたデータであり、
前記副次成分信号強度が、蛍光強度であることを特徴とする、請求項 1 ～ 18 の何れか一項に記載の方法。

【請求項 23】

前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであり、前記工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 が、それぞれ工程 A₁-1、工程 A₁-2、工程 A₁-3-1 及び工程 A₁-4-1 であることを特徴とする、請求項 1 ～ 11 の何れか一項に記載の方法。

[工程 A₁-1]

母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸と、を含む循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデー

10

20

30

40

50

タセット（但し、前記信号の真偽は既知である）を用意する工程。

〔工程 A 1-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A 1）及び前記（A 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

〔工程 A 1-3-1〕

前記工程 A 1-2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

10

（但し、前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで異型であるアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を真とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで同型であるアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を偽とし、

20

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を真とする。）

〔工程 A 1-4-1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

〔請求項 2 4〕

前記主要寄与体が健常人であり、前記副次寄与体が癌細胞であり、前記工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 が、それぞれ工程 A 2-1、工程 A 2-2、工程 A 2-3-1 及び工程 A 2-4-1 であることを特徴とする請求項 1～10 の何れか一項に記載の方法。

30

〔工程 A 2-1〕

前記健常人に関する遺伝情報を含む主要核酸を含有する健常人より採取した核酸サンプルに、癌に関連する多型座位において癌関連変異が導入された前記多型座位の塩基配列情報を含む複数の核酸断片からなる副次核酸を添加することで人工的に調製した混合核酸サンプルの測定により得られるデータを含むデータセットであり、

前記主要核酸における正常型のアレルの存在を示す信号と、前記副次核酸における前記変異を含むアレルの存在を示す信号と、を含むデータセットを用意する工程。

〔工程 A 2-2〕

40

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A 1）及び前記（A 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

〔工程 A 2-3-1〕

前記工程 A 2-2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

（但し、前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加した場合であって、

50

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は真とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は偽とする。

前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加していない場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は偽とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は真とする。）

〔工程 A 2 - 4 - 1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【請求項 25】

以下の工程 A 2' - 1、工程 A 2' - 2、工程 A 2' - 3 - 1 及び工程 A 2' - 4 - 1 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

〔工程 A 2' - 1〕

健常人に関する遺伝情報を含む主要核酸を含有する健常人より採取した核酸サンプルに、癌に関連する単一の多型座位において癌関連変異が導入された前記単一の多型座位の塩基配列情報を含む複数の核酸断片からなる副次核酸を添加することで人工的に調製した、前記副次核酸の含有割合が互いに異なる複数の混合核酸サンプルの測定により得られるデータを含むデータセットであり、

前記主要核酸における正常型のアレルの存在を示す信号と、前記副次核酸における前記変異を含むアレルの存在を示す信号と、を含むデータセットを用意する工程。

〔工程 A 2' - 2〕

前記データセットに含まれるデータのうち、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された前記単一の多型座位に関する、少なくとも以下の (A 1') 及び (A 2') を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程。

(A 1') 前記副次核酸に由来する前記単一の多型座位のアレルの存在を示す副次成分信号強度。

(A 2') 前記単一の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率

〔工程 A 2 - 3 - 1〕

前記工程 A 2' - 2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

(但し、前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加した場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は真とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は偽とする。

前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加していない場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は偽とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は真とする。）

〔工程 A 2 - 4 - 1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信

10

20

30

40

50

信頼性値を算出するためのモデル関数を求める工程。

【請求項 26】

前記主要寄与体が臓器移植のレシピエントであり、前記副次寄与体が移植臓器であり、前記工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 が、それぞれ工程 A3-1、工程 A3-2、工程 A3-3-1 及び工程 A3-4-1 であることを特徴とする、請求項 1～10 の何れか一項に記載の方法。

〔工程 A3-1〕

レシピエントに関する遺伝情報を含む主要核酸と、移植臓器に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

10

〔工程 A3-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A1）及び前記（A2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

〔工程 A3-3-1〕

前記工程 A3-2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。（但し、レシピエントが有していないアレルであって、かつ、ドナーがホモ接合若しくはヘテロ接合で有しているアレルについて、

20

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を真とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。

レシピエント及びドナーの何れもが有していないアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、前記副次成分信号を偽とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、前記副次成分信号を真とする。）

30

〔工程 A3-4-1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【請求項 27】

モデル関数にその説明変数を入力することで、信頼性値を算出する信頼性値の算出方法であり、

前記モデル関数が、

請求項 1～26 の何れか一項に記載の方法で求めた前記モデル関数、

以下の式 1～3 の何れかのモデル関数、又は

40

以下の式 1～3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数であり、

前記説明変数が、以下の工程 B-1 で用意するデータセットに含まれる以下の（B1）、（B2）及び以下の工程 B-2 で求めた合成変数から選ばれる 1 又は 2 以上の数値であることを特徴とする、信頼性値の算出方法。

〔工程 B-1〕

主要寄与体に関する遺伝情報を含む主要核酸を含み、副次寄与体に関する遺伝情報を含む副次核酸を含む若しくは含み得る混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存

50

在を示す信号を含むデータセットを用意する工程。

〔工程 B－2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の（B 1）及び（B 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

（B 1）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

（B 2）前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

10

〔数 1〕

〔式 1〕

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

（ x1: 第 1 主成分、 A1: 遷移領域の勾配、 x01: 半値点）

20

（ただし、式 1 中、A 1 は 1 5 . 4 ～ 1 5 . 6、x 0 1 は - 0 . 8 ～ - 0 . 6 である。）

〔数 2〕

〔式 2〕

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

（ x2: 副次成分信号強度、 A2: 遷移領域の勾配、 x02: 半値点）

30

（ただし、式 2 中、A 2 は 1 . 8 ～ 2 . 0 であり、x 0 2 は 2 . 5 ～ 2 . 7 である。）

〔数 3〕

〔式 3〕

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

（ x3: 副次成分混入率、 A3: 遷移領域の勾配、 x03: 半値点）

40

（ただし、式 3 中、A 3 は 9 . 3 ～ 9 . 5 であり、x 0 3 は 0 . 5 ～ 0 . 7 である。）

〔請求項 2 8〕

前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであり、前記工程 B－1 及び工程 B－2 が、それぞれ工程 B 1－1 及び工程 B 1－2 であることを特徴とする、請求項 2 7 に記載の方法。

〔工程 B 1－1〕

母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸と、を含む循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及

50

び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

【工程 B 1－2】

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記母親においてホモ接合であり、かつ、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B 1）及び前記（B 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

【請求項 29】

前記複数の多型座位が、ヒト個人識別で用いられる多型座位であり、非侵襲的出生前親子鑑定のために信頼性値を算出する方法であることを特徴とする、請求項 28 に記載の方法。

【請求項 30】

前記主要寄与体が検査対象者であり、前記副次寄与体が癌細胞であり、前記混合核酸サンプルが前記検査対象者から採取された循環無細胞核酸サンプルであり、前記工程 B－1 及び工程 B－2 が、それぞれ工程 B 2－1 及び工程 B 2－2 であることを特徴とする、請求項 27 に記載の方法。

【工程 B 2－1】

検査対象者に関する遺伝情報を含む主要核酸を含み、癌細胞に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、癌に関連する複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

【工程 B 2－2】

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、正常型のアレルの存在を示す信号と、変異型のアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B 1）及び前記（B 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

【請求項 31】

前記工程 B 2－2 において、前記データセットに含まれるデータから、前記複数の多型座位の中で、検査対象者において変異型のアレルをホモ接合又はヘテロ接合として有している多型座位に関するデータを除外し、

除外後に残った前記データセットに含まれるデータのうち、前記複数の多型座位の中で、正常型のアレルの存在を示す信号と、変異型のアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B 1）及び前記（B 2）を含む数値群を線形結合させ、1つ以上の合成変数を生成することを特徴とする、請求項 30 に記載の方法。

【請求項 32】

前記主要寄与体が臓器移植のレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルであり、前記工程 B－1 及び工程 B－2 が、それぞれ工程 B 3－1 及び工程 B 3－2 であることを特徴とする、請求項 27 に記載の方法。

【工程 B 3－1】

レシピエントに関する遺伝情報を含む主要核酸を含み、移植臓器に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

【工程 B 3－2】

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B 1）及び前記（B 2）

10

20

30

40

50

を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

【請求項33】

前記複数の多型座位が、ヒト個人識別で用いられる多型座位であり、
移植臓器の定着のモニタリングのために信頼性値を算出する方法であることを特徴とする、請求項32に記載の方法。

【請求項34】

請求項27～33の何れか一項に記載の方法による信頼性値の算出に適さないデータを除外するための除外条件を設定する方法であり、以下の工程C-1-1、工程C-2-1、工程C-3-1及び工程C-4-1を備えることを特徴とする除外条件の設定方法。

【工程C-1-1】

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。（但し、前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルである、又は

前記主要寄与体がレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。）

【工程C-2-1】

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも以下の（C1）、（C2）及び（C3）を含む数値群を線形結合させて得られる合成変数のうち、最も寄与率が高い合成変数を生成する工程。

（C1）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

（C2）前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

（C3）前記特定の多型座位のアレルに起因する信号強度総和から、前記主要核酸に由来する特定の多型座位の一のアレルの存在を示す主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

【工程C-3-1】

前記工程C-2-1における前記線形結合により得られた前記合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程。

【工程C-4-1】

信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件C1として設定する工程。

（除外条件C1）

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも前記（C1）、前記（C2）及び（C3）を含む数値群を線形結合させて得られた、最も寄与率が高い合成変数が、前記工程C-3-1で設定した前記閾値未満に該当するデータセットは除去する。

【請求項35】

請求項 27～33 の何れか一項に記載の方法による信頼性値の算出に適さないデータを除外するための除外条件を設定する方法であり、以下の工程 C-1-2 及び工程 C-2-2、工程 C-3-2 及び工程 C-4-2 を備えることを特徴とする除外条件の設定方法。

〔工程 C-1-2〕

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

（但し、前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであるか、又は

前記主要寄与体がレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。）

〔工程 C-2-2〕

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも以下の（C1）、（C2）及び（C3）を含む数値群を線形結合させて得られる合成変数のうち、一番目又は二番目に寄与率が高い合成変数を生成する工程。

（C1）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

（C2）前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

（C3）前記特定の多型座位のアレルに起因する信号強度総和から、前記主要核酸に由来する特定の多型座位の一のアレルの存在を示す主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

〔工程 C-3-2〕

前記工程 C-2-2 における前記線形結合により得られた前記合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程。

〔工程 C-4-2〕

信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件 C2 として設定する工程。

（除外条件 C2）

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも前記（C1）、前記（C2）及び前記（C3）を含む数値群を線形結合させて得られた、一番目又は二番目に寄与率が高い合成変数が、前記工程 C-3-2 で設定した前記閾値未満に該当するデータセットは除去する。

【請求項 36】

前記多型座位が、ヒト個人識別で用いられる一塩基多型座位であることを特徴とする、請求項 34 又は 35 に記載の方法。

【請求項 37】

移植臓器の定着のモニタリングのための方法であることを特徴とする、請求項 34～36 の何れか一項に記載の方法。

10

20

30

40

50

【請求項 38】

前記外れ値は、請求項 27～33 の何れか一項に記載の方法によって信頼性値を算出したときに、

前記副次核酸に由来する特定のアレルが前記混合核酸サンプルに含まれているにも関わらず、当該アレルの存在を示す信号の信頼性値が 0.8 未満として算出されてしまう場合における当該アレルに関する数値、及び／又は

前記副次核酸に由来する特定のアレルが前記混合核酸サンプルに含まれていないにも関わらず、当該アレルの存在を示す信号の信頼性値が 0.2 以上として算出されてしまう場合における当該アレルに関する数値であることを特徴とする、請求項 34～37 の何れか一項に記載の方法。

10

【請求項 39】

前記外れ値は、前記合成変数の平均値から、その標準偏差の 2 倍以上離れた数値であることを特徴とする、請求項 34～38 の何れか一項に記載の方法。

【請求項 40】

前記工程 B-1 において、

請求項 34 に記載の方法で特定した除外条件 C1 及び／又は請求項 35 に記載の方法で特定した除外条件 C2 に該当するデータセットを除去した後に残ったデータセットを用意することを特徴とする、請求項 32 又は 33 に記載の方法。

【請求項 41】

モデル関数にその説明変数を入力することで、信頼性値を算出する信頼性値の算出方法であり、

前記モデル関数が、

請求項 1～26 の何れか一項に記載の方法で求めた前記モデル関数、

以下の式 1～3 の何れかのモデル関数、又は

以下の式 1～3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数であり、

前記説明変数が、以下の工程 B4-1 で用意するデータセットに含まれる以下の (B1)、(B2) 及び以下の工程 B4-2 で求めた合成変数から選ばれる 1 又は 2 以上の数値であることを特徴とする、信頼性値の算出方法。

20

30

【工程 B4-1】

母親に関する遺伝情報を含む主要核酸と、前記母親の胎内にいる胎児に関する遺伝情報を含む副次核酸と、が含まれる、前記母親から採取された循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、疾患に関連する複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

【工程 B4-2】

前記データセットに含まれるデータから、前記複数の多型座位の中で、母親において変異型のアリルをヘテロ接合として有している多型座位に関するデータを除外し、

除外後に残った前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の (B1) 及び (B2) を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程。

40

(B1) 前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

(B2) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

【数 1】

50

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

(x1:第 1 主成分、 A1:遷移領域の勾配、 x01:半値点)

(ただし、式 1 中、A 1 は 1 5 . 4 ~ 1 5 . 6、x 0 1 は - 0 . 8 ~ - 0 . 6 である。)

10

【数 2】

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

(x2:副次成分信号強度、 A2:遷移領域の勾配、 x02:半値点)

20

(ただし、式 2 中、A 2 は 1 . 8 ~ 2 . 0 であり、x 0 2 は 2 . 5 ~ 2 . 7 である。)

【数 3】

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

(x3:副次成分混入率、 A3:遷移領域の勾配、 x03:半値点)

30

(ただし、式 3 中、A 3 は 9 . 3 ~ 9 . 5 であり、x 0 3 は 0 . 5 ~ 0 . 7 である。)

【請求項 4 2】

疾患リスクを評価する非侵襲的出生前検査のために信頼性値を算出する方法であることを特徴とする、請求項 4 1 に記載の方法。

【請求項 4 3】

請求項 1 ~ 4 2 の何れか一項に記載の方法をコンピュータに実行させるためのプログラム。

【請求項 4 4】

請求項 4 3 に記載のプログラムを記録した記録媒体。

【請求項 4 5】

請求項 1 ~ 2 6 の何れか一項に記載の方法で作成されたモデル関数、
以下の式 1 ~ 3 の何れかのモデル関数、又は
以下の式 1 ~ 3 で表されるモデル関数からなる群から選択される
2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数
が記録された請求項 4 4 に記載の記憶媒体。

40

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

($x1$: 第 1 主成分、 $A1$: 遷移領域の勾配、 $x01$: 半値点)

(ただし、式 1 中、 $A1$ は 15.4 ~ 15.6、 $x01$ は -0.8 ~ -0.6 である。)

10

【数 2】

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

($x2$: 副次成分信号強度、 $A2$: 遷移領域の勾配、 $x02$: 半値点)

20

(ただし、式 2 中、 $A2$ は 1.8 ~ 2.0 であり、 $x02$ は 2.5 ~ 2.7 である。)

【数 3】

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3$: 遷移領域の勾配、 $x03$: 半値点)

30

(ただし、式 3 中、 $A3$ は 9.3 ~ 9.5 であり、 $x03$ は 0.5 ~ 0.7 である。)

【請求項 46】

請求項 1 ~ 26 の何れか一項に記載の方法で作成されたモデル関数、

以下の式 1 ~ 3 の何れかのモデル関数、又は

以下の式 1 ~ 3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数

が記録された記憶部と、請求項 27 ~ 33 及び 40 ~ 42 の何れか一項に記載の方法を実行する処理部と、を備える、信頼性値算出システム。

40

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

($x1$: 第 1 主成分、 $A1$: 遷移領域の勾配、 $x01$: 半値点)

50

(ただし、式 1 中、 A_1 は 15.4 ～ 15.6、 x_{01} は -0.8 ～ -0.6 である。)

【数 2】

[式 2]

$$f_2(x_2) = \frac{1}{1 + e^{-A_2(x_2 - x_{02})}}$$

(x_2 : 副次成分信号強度、 A_2 : 遷移領域の勾配、 x_{02} : 半値点)

10

(ただし、式 2 中、 A_2 は 1.8 ～ 2.0 であり、 x_{02} は 2.5 ～ 2.7 である。)

【数 3】

[式 3]

$$f_3(x_3) = \frac{1}{1 + e^{-A_3(x_3 - x_{03})}}$$

(x_3 : 副次成分混入率、 A_3 : 遷移領域の勾配、 x_{03} : 半値点)

20

(ただし、式 3 中、 A_3 は 9.3 ～ 9.5 であり、 x_{03} は 0.5 ～ 0.7 である。)

【請求項 47】

前記記憶部に、請求項 34 に記載の方法で作成された除外条件 C1 及び／又は請求項 35 に記載の方法で作成された除外条件 C2 が記録されており、

前記処理部は、請求項 40 に記載の方法を実行することを特徴とする、請求項 46 に記載の信頼性値算出システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は SNPs などの解析データのデータ処理に関する。

【背景技術】

【0002】

親子関係の存在・不存在が不明瞭であることは法律的、家族関係などに大きな影響を及ぼす。妊娠している女性の胎内にいる胎児の実父が誰であるのか確信が持てない場合、正しい実父を決定するいくつかの方法がある。

1つの方法としては、出産まで待ち、子と擬父のゲノムDNAを解析してこれらと比較する方法が挙げられる。しかし、子の出生前にその実父を知りたいというニーズは多い。出生前に親子関係を鑑別する方法としては、絨毛診断や羊水穿刺によって回収した遺伝物質を解析する方法が挙げられるが、これらは侵襲性であり、流産リスクがあるという問題がある。

40

【0003】

上述した侵襲性の診断方法の問題に鑑み、血液に混入した循環無細胞DNA (cell-free DNA, cfDNA) を解析する方法を親子鑑定に応用することが行われている。母親の血液循環に混入した胎児由来の遺伝物質である胎児循環無細胞DNA (cell-free fetal DNA, cffDNA) の分析を行うことにより、非侵襲的出生前親子鑑定 (Non-Invasive Prenatal Paternity Test, NIPT) を実施することが可能となる (例えば特許文献1)。

【0004】

ただし cffDNA の分析には解決しなければならない重要な課題がある。母体血には

50

胎児由来の遺伝物質の他、母親由来の遺伝物質も当然に含まれており、後者の方が前者に比して圧倒的に量が多い。したがって、c f DNAの解析により得られる胎児の遺伝物質の存在を示す信号は微弱なものとなる。そのため、c f DNAの解析により得られた胎児の遺伝物質の存在を示唆する信号が、本当に胎児の遺伝物質に由来するものなのか、それともノイズであるのかを判別することは極めて困難である。

【0005】

また、c f DNAの分析技術の応用分野としては上記の出生前遺伝学的検査の他、癌スクリーニング検査や抗癌治療の経過の評価検査などを含む癌検査が挙げられる。癌細胞が免疫によって破壊されたり、自ら細胞死（アポトーシス）を起こしたり、また、血中を循環する循環腫瘍細胞（CTC）が何らかの影響によって血中で破壊されたりすると、癌細胞のゲノムDNAが血中に漏出することになる。この癌細胞に由来するc f DNAは特別にc t DNA（circulating tumor DNA）と呼ばれることもある。一方で癌細胞のゲノムDNAには、特異的な一塩基変異が発生することが知られている。癌に関連する変異が生じる多型座位の配列解析をすることによって、癌の早期発見が可能になるものと期待されている（例えば特許文献2）。

これらの知見を組合せ、血中を循環するc f DNAを抽出し、癌に関連する変異が生じる多型座位を解析することで癌検査が可能となる。

ただ、c f DNAの大部分は検査対象者自身の正常型のゲノムDNAに由来し、癌細胞由来のc f DNAが含まれる割合は極微小である。そのため、上記の出生前遺伝学的検査と同様に、c f DNAの解析により得られた癌に関連する変異の存在を示唆する信号が、本当に癌細胞のゲノムDNAに由来するものなのか、それともノイズであるのかを判別することが極めて困難であるという問題がある。

【0006】

さらに、c f DNAの分析技術の応用分野として移植臓器の定着のモニタリングなどが挙げられる。免疫抑制剤などの改善により成功率の向上は見られるものの、拒絶反応の問題は未だ移植臓器の長期定着にとって大きな問題である。拒絶反応により移植臓器が障害されたり、壊死に至ったりすると、移植臓器を構成する細胞から血中にゲノムDNAが漏出する。この移植臓器由来のc f DNA（特別にd d c f DNAと呼ばれることもある）が移植臓器障害のバイオマーカーとして期待されている。具体的には、ドナーとレシピエントの個人識別が可能な一塩基置換（SNPs）を選定し、次世代シーケンサーなどを利用してレシピエントの血液中に漏出する極微量のd d c f DNAの定量を行う方法である（例えば特許文献3）。

ただ、c f DNAの大部分はレシピエントのゲノムDNAに由来し、d d c f DNAが含まれる割合は極微小であるため、上記の出生前遺伝学的検査と同様に、c f DNAの解析により得られたd d c f DNAの存在を示唆する信号が、本当に移植臓器のゲノムDNAに由来するものなのか、それともノイズであるのかを判別することが極めて困難であるという問題がある。

【先行技術文献】

【特許文献】

【0007】

【文献】特表2014-502845号公報

再表2017-094805号公報

特表2020-529648号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

本発明の解決すべき課題はc f f DNA、c t DNA、d d c f DNAのような副次核酸を微小な割合で含む混合核酸サンプルの分析データにおける、副次核酸の存在を示す信号の信頼性を評価する新規の技術を提供することにある。

【課題を解決するための手段】

【0009】

上記課題を解決する本発明は以下の通りである。

【0010】

〔1〕 以下の工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

〔工程 A-1〕

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

10

〔工程 A-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の（A1）及び（A2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

（A1）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度

（A2）前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率

〔工程 A-3-1〕

20

前記工程 A-2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

〔工程 A-4-1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【0011】

〔2〕 前記合成変数が、少なくとも前記（A1）及び前記（A2）を含む数値群について主成分分析を行うことで生成し得る合成変数であることを特徴とする、〔1〕に記載の方法。

30

【0012】

〔3〕 前記工程 A-3-1 及び工程 A-4-1 においてモデル関数の作成のために用いる前記合成変数が、工程 A-2 において生成した1以上の合成変数のうち、最も寄与率が高い合成変数であることを特徴とする、〔2〕に記載の方法。

【0013】

〔4〕 前記工程 A-2 が、少なくとも前記（A1）及び前記（A2）を含む数値群について主成分分析を行い、1つ以上の主成分を合成変数として生成する工程であることを特徴とする、〔1〕～〔3〕の何れかに記載の方法。

【0014】

40

〔5〕 前記工程 A-2 が、前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A1）及び前記（A2）を含み、さらに以下の（A3）～（A5）から選ばれる1又は2以上を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程であることを特徴とする、〔1〕～〔4〕の何れかに記載の方法。

（A3）前記主要核酸に由来する、特定の多型座位の一のアレルの存在を示す主要成分信号強度。

（A4）前記特定の多型座位のアレルに起因する信号強度総和に対する前記主要成分信号強度の割合である、主要成分混入率。

50

(A 5) 前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

【0015】

〔6〕 前記工程 A-2 が、前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A 1) 及び (A 2) を含み、さらに以下の (A 3) ~ (A 5) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程であることを特徴とする、〔1〕 ~ 〔5〕の何れかに記載に記載の方法。

(A 3) 前記主要核酸に由来する、特定の多型座位の一のアレルの存在を示す主要成分信号強度。

10

(A 4) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記主要成分信号強度の割合である、主要成分混入率。

(A 5) 前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

【0016】

〔7〕 前記回帰分析が最小二乗法であることを特徴とする、〔1〕 ~ 〔6〕の何れかに記載の方法。

【0017】

〔8〕 前記合成変数を表す1次斉次多項式において、副次成分信号強度又は副次成分混入率が最大に重みづけされることを特徴とする、〔1〕 ~ 〔7〕の何れかに記載の方法。

20

【0018】

〔9〕 前記工程 A-2 において線形結合する数値群に含まれる数値は、標準化された数値であることを特徴とする、〔1〕 ~ 〔8〕の何れかに記載の方法。

【0019】

〔10〕 前記工程 A-2 において、2つ以上の合成変数を生成し、
前記工程 A-3-1 において、前記2つ以上の合成変数のそれぞれについて、信頼性値の付与を行い、

前記工程 A-4-1 において、前記2つ以上の合成変数のそれぞれを説明変数とする互いに独立した2以上のモデル関数を作成し、

30

さらに、前記2以上のモデル関数を互いに乗じ、乗算で表されるモデル関数を作成する工程を備えることを特徴とする、〔1〕 ~ 〔9〕の何れかに記載の方法。

【0020】

〔11〕 以下の工程 A-1、工程 A-3-2 及び工程 A-4-2 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

〔工程 A-1〕

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

40

〔工程 A-3-2〕

前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度を複数の区分し、各区分に含まれる前記副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記副次成分信号強度に対応する確率として付与する工程。

〔工程 A-4-2〕

前記各区分に含まれる前記副次成分信号強度と、前記各区分に含まれる前記副次成分信号強度に対応する確率について回帰分析を行い、前記副次成分信号強度を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

50

【0021】

【12】 以下の工程A-1、工程A-3-3及び工程A-4-3を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

【工程A-1】

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

【工程A-3-3】

前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合である副次成分混入率を複数の区分し、各区分に含まれる前記副次成分混入率に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記副次成分混入率に対応する確率として付与する工程。

【工程A-4-3】

前記各区分に含まれる前記副次成分混入率と、前記各区分に含まれる前記副次成分混入率に対応する確率について回帰分析を行い、前記副次成分混入率を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【0022】

【13】 前記モデル関数が、シグモイド関数であることを特徴とする、【1】～【12】の何れかに記載の方法。

【0023】

【14】 前記モデル関数が、2個の媒介変数を有するシグモイド関数であることを特徴とする、【1】～【13】の何れかに記載の方法。

【0024】

【15】 【1】～【10】の何れかに記載の方法で作成したモデル関数と、

【11】に記載の方法で作成したモデル関数と、

【12】に記載の方法で作成したモデル関数と、

からなる群から選択される2以上のモデル関数を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【0025】

【16】 【1】～【10】の何れかに記載の方法で作成したモデル関数と、

【11】に記載の方法で作成したモデル関数、及び／又は、【12】に記載の方法で作成したモデル関数と、

を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【0026】

【17】 【1】～【10】の何れかに記載の方法で作成したモデル関数と、

【11】に記載の方法で作成したモデル関数と、

【12】に記載の方法で作成したモデル関数と、

を互いに乗じ、乗算で表されるモデル関数を作成する工程を備える、モデル関数の作成方法。

【0027】

【18】 前記多型座位が、一塩基多型を含む座位であることを特徴とする、【1】～【17】の何れかに記載の方法。

【0028】

【19】 前記データセットが、塩基配列解析、デジタルPCR、マイクロアレイ、リアルタイムPCR又は質量分析により取得されたデータであることを特徴とする、【1】～【18】の何れかに記載の方法。

10

20

30

40

50

【0029】

【20】 前記データセットが塩基配列解析により取得されたデータであり、前記副次成分信号強度が、配列タグのカウント数、リード数、イオン濃度又は電気信号であることを特徴とする、【1】～【18】の何れかに記載の方法。

【0030】

【21】 前記データセットがデジタルPCRにより取得されたデータであり、前記副次成分信号強度が、蛍光が観察されたウェル数であることを特徴とする、【1】～【18】の何れかに記載の方法。

【0031】

【22】 前記データセットがマイクロアレイにより取得されたデータであり、前記副次成分信号強度が、蛍光強度であることを特徴とする、【1】～【18】の何れかに記載の方法。

10

【0032】

【23】 前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであり、前記工程A-1、工程A-2、工程A-3-1及び工程A-4-1が、それぞれ工程A₁-1、工程A₁-2、工程A₁-3-1及び工程A₁-4-1であることを特徴とする、【1】～【11】の何れかに記載の方法。

【工程A₁-1】

母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸と、を含む循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

20

【工程A₁-2】

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A₁）及び前記（A₂）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

【工程A₁-3-1】

前記工程A₁-2で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

30

（但し、前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで異型であるアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を真とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで同型であるアレルについて、

40

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を偽とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を真とする。）

【工程A₁-4-1】

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【0033】

50

〔24〕 前記主要寄与体が健常人であり、前記副次寄与体が癌細胞であり、前記工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 が、それぞれ工程 A₂-1、工程 A₂-2、工程 A₂-3-1 及び工程 A₂-4-1 であることを特徴とする〔1〕～〔10〕の何れかに記載の方法。

〔工程 A₂-1〕

前記健常人に関する遺伝情報を含む主要核酸を含有する健常人より採取した核酸サンプルに、癌に関連する多型座位において癌関連変異が導入された前記多型座位の塩基配列情報を含む複数の核酸断片からなる副次核酸を添加することで人工的に調製した混合核酸サンプルの測定により得られるデータを含むデータセットであり、

前記主要核酸における正常型のアレルの存在を示す信号と、前記副次核酸における前記変異を含むアレルの存在を示す信号と、を含むデータセットを用意する工程。

〔工程 A₂-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A1）及び前記（A2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

〔工程 A₂-3-1〕

前記工程 A₂-2 で生成された合成変数を複数の区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

（但し、前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加した場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は真とし、

該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は偽とする。

前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加していない場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は偽とし、

該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は真とする。）

〔工程 A₂-4-1〕

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

〔0034〕

〔25〕 以下の工程 A₂'-1、工程 A₂'-2、工程 A₂'-3-1 及び工程 A₂'-4-1 を含むことを特徴とする、副次成分信号の信頼性値を算出するためのモデル関数の作成方法。

〔工程 A₂'-1〕

健常人に関する遺伝情報を含む主要核酸を含有する健常人より採取した核酸サンプルに、癌に関連する単一の多型座位において癌関連変異が導入された前記単一の多型座位の塩基配列情報を含む複数の核酸断片からなる副次核酸を添加することで人工的に調製した、前記副次核酸の含有割合が互いに異なる複数の混合核酸サンプルの測定により得られるデータを含むデータセットであり、

前記主要核酸における正常型のアレルの存在を示す信号と、前記副次核酸における前記変異を含むアレルの存在を示す信号と、を含むデータセットを用意する工程。

〔工程 A₂'-2〕

前記データセットに含まれるデータのうち、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された前記単一の多型座位に関する、少なくとも以下の（A1'）及び（A2'）を含む数値群を

線形結合させ、1つ以上の合成変数を生成する工程。

(A 1') 前記副次核酸に由来する前記単一の多型座位のアレルの存在を示す副次成分信号強度。

(A 2') 前記単一の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率

[工程 A 2-3-1]

前記工程 A 2'-2 で生成された合成変数を複数に区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。

(但し、前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加した場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は真とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は偽とする。

前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加していない場合であって、

該核酸断片について副次成分信号が検出された場合には、当該副次成分信号は偽とし、
該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は真とする。)

[工程 A 2-4-1]

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【0035】

[26] 前記主要寄与体が臓器移植のレシピエントであり、前記副次寄与体が移植臓器であり、前記工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 が、それぞれ工程 A 3-1、工程 A 3-2、工程 A 3-3-1 及び工程 A 3-4-1 であることを特徴とする、[1] ~ [10] の何れかに記載の方法。

[工程 A 3-1]

レシピエントに関する遺伝情報を含む主要核酸と、移植臓器に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット (但し、前記信号の真偽は既知である) を用意する工程。

[工程 A 3-2]

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A 1) 及び前記 (A 2) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

[工程 A 3-3-1]

前記工程 A 3-2 で生成された合成変数を複数に区分し、各区分に含まれる前記合成変数に対応する副次成分信号強度のうちこれが真であった割合を各区分に含まれる前記合成変数に対応する確率として付与する工程。(但し、レシピエントが有していないアレルであって、かつ、ドナーがホモ接合若しくはヘテロ接合で有しているアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、当該副次成分信号を真とし、

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。

レシピエント及びドナーの何れもが有していないアレルについて、

主要成分信号と区別されて前記副次成分信号が検出された場合、前記副次成分信号を偽とし、

10

20

30

40

50

主要成分信号と区別されて前記副次成分信号が検出されなかった場合、前記副次成分信号を真とする。)

[工程 A 3-4-1]

前記各区分に含まれる前記合成変数と、前記各区分に含まれる前記合成変数に対応する確率について回帰分析を行い、前記合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める工程。

【0036】

[27] モデル関数にその説明変数を入力することで、信頼性値を算出する信頼性値の算出方法であり、

前記モデル関数が、

[1] ~ [26] の何れかに記載の方法で求めた前記モデル関数、

以下の式 1 ~ 3 の何れかのモデル関数、又は

以下の式 1 ~ 3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数であり、

前記説明変数が、以下の工程 B-1 で用意するデータセットに含まれる以下の (B1)、(B2) 及び以下の工程 B-2 で求めた合成変数から選ばれる 1 又は 2 以上の数値であることを特徴とする、信頼性値の算出方法。

[工程 B-1]

主要寄与体に関する遺伝情報を含む主要核酸を含み、副次寄与体に関する遺伝情報を含む副次核酸を含む若しくは含み得る混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

[工程 B-2]

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の (B1) 及び (B2) を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程。

(B1) 前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

(B2) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

(x1: 第 1 主成分、 A1: 遷移領域の勾配、 x01: 半値点)

(ただし、式 1 中、A1 は 15.4 ~ 15.6、x01 は -0.8 ~ -0.6 である。)

【数 2】

10

20

30

40

50

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

($x2$: 副次成分信号強度、 $A2$: 遷移領域の勾配、 $x02$: 半値点)

(ただし、式 2 中、 $A2$ は 1.8 ~ 2.0 であり、 $x02$ は 2.5 ~ 2.7 である。)

10

【数 3】

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3$: 遷移領域の勾配、 $x03$: 半値点)

(ただし、式 3 中、 $A3$ は 9.3 ~ 9.5 であり、 $x03$ は 0.5 ~ 0.7 である。)

20

【0037】

〔28〕 前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであり、前記工程 B-1 及び工程 B-2 が、それぞれ工程 B₁-1 及び工程 B₁-2 であることを特徴とする、〔27〕に記載の方法。

〔工程 B₁-1〕

母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸と、を含む循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

30

〔工程 B₁-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記母親においてホモ接合であり、かつ、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (B1) 及び前記 (B2) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

【0038】

〔29〕 前記複数の多型座位が、ヒト個人識別で用いられる多型座位であり、

非侵襲的出生前親子鑑定のために信頼性値を算出する方法であることを特徴とする、〔28〕に記載の方法。

40

【0039】

〔30〕 前記主要寄与体が検査対象者であり、前記副次寄与体が癌細胞であり、前記混合核酸サンプルが前記検査対象者から採取された循環無細胞核酸サンプルであり、前記工程 B-1 及び工程 B-2 が、それぞれ工程 B₂-1 及び工程 B₂-2 であることを特徴とする、〔27〕に記載の方法。

〔工程 B₂-1〕

検査対象者に関する遺伝情報を含む主要核酸を含み、癌細胞に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、癌に関連する複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

50

[工程 B 2-2]

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、正常型のアレルの存在を示す信号と、変異型のアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (B 1) 及び前記 (B 2) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

[0040]

[31] 前記工程 B 2-2 において、前記データセットに含まれるデータから、前記複数の多型座位の中で、検査対象者において変異型のアレルをホモ接合又はヘテロ接合として有している多型座位に関するデータを除外し、

除外後に残った前記データセットに含まれるデータのうち、前記複数の多型座位の中で、正常型のアレルの存在を示す信号と、変異型のアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (B 1) 及び前記 (B 2) を含む数値群を線形結合させ、1つ以上の合成変数を生成することを特徴とする、[30]に記載の方法。

[0041]

[32] 前記主要寄与体が臓器移植のレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルであり、前記工程 B-1 及び工程 B-2 が、それぞれ工程 B 3-1 及び工程 B 3-2 であることを特徴とする、[27]に記載の方法。

[工程 B 3-1]

レシピエントに関する遺伝情報を含む主要核酸を含み、移植臓器に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

[工程 B 3-2]

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (B 1) 及び前記 (B 2) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程。

[0042]

[33] 前記複数の多型座位が、ヒト個人識別で用いられる多型座位であり、移植臓器の定着のモニタリングのために信頼性値を算出する方法であることを特徴とする、[32]に記載の方法。

[0043]

[34] [27] ~ [33] の何れかに記載の方法による信頼性値の算出に適さないデータを除外するための除外条件を設定する方法であり、以下の工程 C-1-1、工程 C-2-1、工程 C-3-1 及び工程 C-4-1 を備えることを特徴とする除外条件の設定方法。

[工程 C-1-1]

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット (但し、前記信号の真偽は既知である) を用意する工程。 (但し、前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルである、又は

前記主要寄与体がレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。)

[工程 C-2-1]

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合で

10

20

30

40

50

あり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも以下の（Ｃ１）、（Ｃ２）及び（Ｃ３）を含む数値群を線形結合させて得られる合成変数のうち、最も寄与率が高い合成変数を生成する工程。

（Ｃ１）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

（Ｃ２）前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

（Ｃ３）前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

〔工程Ｃ－３－１〕

前記工程Ｃ－２－１における前記線形結合により得られた前記合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程。

〔工程Ｃ－４－１〕

信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件Ｃ１として設定する工程。

（除外条件Ｃ１）

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも前記（Ｃ１）、前記（Ｃ２）及び前記（Ｃ３）を含む数値群を線形結合させて得られた、最も寄与率が高い合成変数が、前記工程Ｃ－３－１で設定した前記閾値未満に該当するデータセットは除去する。

〔００４４〕

〔３５〕〔２７〕～〔３３〕の何れかに記載の方法による信頼性値の算出に適さないデータを除外するための除外条件を設定する方法であり、以下の工程Ｃ－１－２及び工程Ｃ－２－２、工程Ｃ－３－２及び工程Ｃ－４－２を備えることを特徴とする除外条件の設定方法。

〔工程Ｃ－１－２〕

主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸と、が含まれる、混合核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセット（但し、前記信号の真偽は既知である）を用意する工程。

（但し、前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルであるか、又は

前記主要寄与体がレシピエントであり、前記副次寄与体が移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。）

〔工程Ｃ－２－２〕

前記母親においてホモ接合であり、父親においてホモ接合であり、かつ、前記母親と前記父親とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも以下の（Ｃ１）、（Ｃ２）及び（Ｃ３）を含む数値群を線形結合させて得られる合成変数のうち、一番目又は二番目に寄与率が高い合成変数を生成する工程。

（Ｃ１）前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

10

20

30

40

50

(C2) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

(C3) 前記特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるノイズ。

[工程C-3-2]

前記工程C-2-2における前記線形結合により得られた前記合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程。

[工程C-4-2]

信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件C2として設定する工程。

(除外条件C2)

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸と、が含まれる混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも前記(C1)、前記(C2)及び前記(C3)を含む数値群を線形結合させて得られた、一番目又は二番目に寄与率が高い合成変数が、前記工程C-3-2で設定した前記閾値未満に該当するデータセットは除去する。

[0045]

[36] 前記多型座位が、ヒト個人識別で用いられる一塩基多型座位であることを特徴とする、[34又は35]に記載の方法。

[0046]

[37] 移植臓器の定着のモニタリングのための方法であることを特徴とする、[34]～[36]の何れかに記載の方法。

[0047]

[38] 前記外れ値は、[27]～[33]の何れかに記載の方法によって信頼性値を算出したときに、

前記副次核酸に由来する特定のアレルが前記混合核酸サンプルに含まれているにも関わらず、当該アレルの存在を示す信号の信頼性値が0.8未満として算出されてしまう場合における当該アレルに関する数値、及び／又は

前記副次核酸に由来する特定のアレルが前記混合核酸サンプルに含まれていないにも関わらず、当該アレルの存在を示す信号の信頼性値が0.2以上として算出されてしまう場合における当該アレルに関する数値であることを特徴とする、[34]～[37]の何れかに記載の方法。

[0048]

[39] 前記外れ値は、前記合成変数の平均値から、その標準偏差の2倍以上離れた数値であることを特徴とする、[34]～[38]の何れかに記載の方法。

[0049]

[40] 前記工程B-1において、

[34]に記載の方法で特定した除外条件C1及び／又は[35]に記載の方法で特定した除外条件C2に該当するデータセットを除去した後に残ったデータセットを用意することを特徴とする、[32]又は[33]に記載の方法。

[0050]

[41] モデル関数にその説明変数を入力することで、信頼性値を算出する信頼性値の算出方法であり、

前記モデル関数が、

[1]～[26]の何れかに記載の方法で求めた前記モデル関数、

10

20

30

40

50

以下の式 1 ～ 3 の何れかのモデル関数、又は

以下の式 1 ～ 3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数であり、

前記説明変数が、以下の工程 B 4 - 1 で用意するデータセットに含まれる以下の (B 1)、(B 2) 及び以下の工程 B 4 - 2 で求めた合成変数から選ばれる 1 又は 2 以上の数値であることを特徴とする、信頼性値の算出方法。

[工程 B 4 - 1]

母親に関する遺伝情報を含む主要核酸と、前記母親の胎内にいる胎児に関する遺伝情報を含む副次核酸と、が含まれる、前記母親から採取された循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、疾患に関連する複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程。

[工程 B 4 - 2]

前記データセットに含まれるデータから、前記複数の多型座位の中で、母親において変異型のアレルをヘテロ接合として有している多型座位に関するデータを除外し、

除外後に残った前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも以下の (B 1) 及び (B 2) を含む数値群を線形結合させ、1 つ以上の合成変数を生成する工程。

(B 1) 前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

(B 2) 前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

(x1: 第 1 主成分、 A1: 遷移領域の勾配、 x01: 半値点)

(ただし、式 1 中、A 1 は 1 5 . 4 ～ 1 5 . 6、x 0 1 は - 0 . 8 ～ - 0 . 6 である。)

【数 2】

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

(x2: 副次成分信号強度、 A2: 遷移領域の勾配、 x02: 半値点)

(ただし、式 2 中、A 2 は 1 . 8 ～ 2 . 0 であり、x 0 2 は 2 . 5 ～ 2 . 7 である。)

【数 3】

10

20

30

40

50

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3$: 遷移領域の勾配、 $x03$: 半値点)

(ただし、式 3 中、 $A3$ は 9.3 ~ 9.5 であり、 $x03$ は 0.5 ~ 0.7 である。)

【0051】

【42】 疾患リスクを評価する非侵襲的出生前検査のために信頼性値を算出する方法であることを特徴とする、【41】に記載の方法。

【0052】

【43】 【1】 ~ 【42】 の何れかに記載の方法をコンピュータに実行させるためのプログラム。

【0053】

【44】 【43】 に記載のプログラムを記録した記録媒体。

【0054】

【45】 【1】 ~ 【26】 の何れかに記載の方法で作成されたモデル関数、

以下の式 1 ~ 3 の何れかのモデル関数、又は

以下の式 1 ~ 3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数

が記録された記憶媒体。

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

($x1$: 第 1 主成分、 $A1$: 遷移領域の勾配、 $x01$: 半値点)

(ただし、式 1 中、 $A1$ は 15.4 ~ 15.6、 $x01$ は -0.8 ~ -0.6 である。)

【数 2】

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

($x2$: 副次成分信号強度、 $A2$: 遷移領域の勾配、 $x02$: 半値点)

(ただし、式 2 中、 $A2$ は 1.8 ~ 2.0 であり、 $x02$ は 2.5 ~ 2.7 である。)

【数 3】

10

20

30

40

50

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3$: 遷移領域の勾配、 $x03$: 半値点)

(ただし、式 3 中、 $A3$ は 9.3 ~ 9.5 であり、 $x03$ は 0.5 ~ 0.7 である。)

【0055】

【46】 [1] ~ [26] の何れかに記載の方法で作成されたモデル関数、

以下の式 1 ~ 3 の何れかのモデル関数、又は

以下の式 1 ~ 3 で表されるモデル関数からなる群から選択される 2 以上のモデル関数を互いに乗じ、乗算で表されるモデル関数

が記録された記憶部と、[27] ~ [33] 及び [40] ~ [42] の何れかに記載の方法を実行する処理部と、を備える、信頼性値算出システム。

【数 1】

[式 1]

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

($x1$: 第 1 主成分、 $A1$: 遷移領域の勾配、 $x01$: 半値点)

(ただし、式 1 中、 $A1$ は 15.4 ~ 15.6、 $x01$ は -0.8 ~ -0.6 である。)

【数 2】

[式 2]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

($x2$: 副次成分信号強度、 $A2$: 遷移領域の勾配、 $x02$: 半値点)

(ただし、式 2 中、 $A2$ は 1.8 ~ 2.0 であり、 $x02$ は 2.5 ~ 2.7 である。)

【数 3】

[式 3]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3$: 遷移領域の勾配、 $x03$: 半値点)

(ただし、式 3 中、 $A3$ は 9.3 ~ 9.5 であり、 $x03$ は 0.5 ~ 0.7 である。)

【0056】

【47】 前記記憶部に、[34] に記載の方法で作成された除外条件 C1 及び／又は [

35]に記載の方法で作成された除外条件C2が記録されており、

前記処理部は、[40]に記載の方法を実行することを特徴とする、[46]に記載の信頼性値算出システム。

【発明の効果】

【0057】

本発明のモデル関数の作成方法によれば、c f f DNA、c t DNA、d d c f DNAのような副次核酸を微小な割合で含む混合核酸サンプルの分析データにおける、副次成分信号の信頼性値を算出するためのモデル関数を作成することができる。

【0058】

また、本発明の信頼性値の算出方法によれば、c f f DNA、c t DNA、d d c f DNAのような副次核酸を微小な割合で含む混合核酸サンプルの分析データにおける、副次成分信号の信頼性値を算出することができる。

【0059】

また、本発明の除外条件の設定方法によれば、前記モデル関数に入力すべき説明変数のデータを絞るため、データセットのうち除外すべきものを判定する除外条件を設定することができる。

【図面の簡単な説明】

【0060】

【図1】モデル関数 $f_1(x_1)$ を示すシグモイド曲線を示す。縦軸の「確率」は信頼性値、横軸の「主成分1」は主成分分析により得られた第1主成分である。図中の白抜きのデータポイントは回帰分析に用いた信頼性値と第1主成分を示す。

【図2】モデル関数 $f_2(x_2)$ を示すシグモイド曲線を示す。縦軸の「確率」は信頼性値、横軸の「胎児マイナーカウント」は副次成分信号強度絶対値である。図中の白抜きのデータポイントは回帰分析に用いた信頼性値と副次成分信号強度絶対値を示す。

【図3】モデル関数 $f_3(x_3)$ を示すシグモイド曲線を示す。縦軸の「確率」は信頼性値、横軸の「胎児マイナーフリークエンシー」は副次成分混入率である。図中の白抜きのデータポイントは回帰分析に用いた信頼性値と副次成分混入率を示す。

【図4】試験例2で算出した信頼性値(Fidelity)の分布図である。左は父母がそれぞれホモ接合で有している互いに異型のSNPsに関する信頼性値を集計したものである。右は父母がそれぞれホモ接合で有している互いに同型のSNPsに関する信頼性値を集計したものである。

【図5】除外条件1の検討のために作成された主成分分析により得られた各主成分をy軸、信頼性値をx軸にプロットした散布図である。左から第1主成分、第2主成分、第3主成分、第4主成分、第5主成分をy軸にとった散布図を示す。

【図6】除外条件2の検討のために作成された主成分分析により得られた各主成分をy軸、信頼性値をx軸にプロットした散布図である。左から第1主成分、第2主成分、第3主成分、第4主成分、第5主成分をy軸にとった散布図を示す。

【図7】試験例4において算出した信頼性値(Fidelity)の分布図である。左は父母がそれぞれホモ接合で有している互いに異型のSNPsに関する信頼性値を集計したものである。右は父母がそれぞれホモ接合で有している互いに同型のSNPsに関する信頼性値を集計したものである。

【図8】試験例5において算出した信頼性値(Fidelity)の分布図である。左は父母がそれぞれホモ接合で有している互いに異型と同型のSNPsに関する信頼性値を数で集計したものである。右はそれぞれ異なるNGSターゲットパネル解析となる試験例2と試験例5で算出した信頼性値(Fidelity)の割合を示したものである。

【図9】試験例6において生まれた子供の解析から確定されたSNPs遺伝型に対する信頼性値(Fidelity)を集計したグラフである。副次成分信号存在に対する真偽を示す父の遺伝型を考慮せず、母ホモSNPs信頼性値(Fidelity)の分布図を数で集計したものである。

【図10】試験例2と試験例8において算出した信頼性値(Fidelity)の分布図

である。左は父母がそれぞれホモ接合で有している互いに異型のSNPsに関する信頼性値を集計したものである（胎児遺伝型の正解はヘテロ接合）。右は父母がそれぞれホモ接合で有している互いに同型のSNPsに関する信頼性値である。

【図11】試験例6と試験例9において算出した信頼性値（Fidelity）の分布図である。左は母がホモ接合で有しており、かつ、新生児がヘテロ接合で有しているSNPsに関する信頼性値を集計したものである。右は母がホモ接合で有しており、かつ、新生児がホモ接合で有しているSNPsに関する信頼性値である。

【発明を実施するための形態】

【0061】

以下、本発明のモデル関数の作成方法、信頼性値の算出方法、除外条件の設定方法の順に具体的な実施形態の説明を加える。なお、本発明の範囲は以下に説明した具体的な実施形態に限定されない。

【0062】

<1>モデル関数の作成方法

以下、本発明のモデル関数の作成方法の実施形態について詳述する。「<1-1>概要」の項目においては、本発明のモデル関数の作成方法の概要説明を加え、「<1-2>出生前遺伝学的検査」の項目においては出生前遺伝学的検査への応用例についての具体的な説明を加え、「<1-3>癌検査」の項目においては癌検査への応用例について具体的な説明を加え、「<1-4>移植臓器の定着のモニタリング」の項目においては、移植臓器の定着のモニタリングへの応用例について具体的な説明を加える。

【0063】

<1-1>概要

本発明のモデル関数の作成方法は、工程A-1、工程A-2、工程A-3-1及び工程A-4-1を必須工程として含む。以下、順に説明する。

【0064】

【工程A-1】

工程A-1は、混合核酸サンプルの測定により得られるデータセットを用意する工程である。

「混合核酸サンプル」とは、複数の寄与体に関する遺伝情報を含むサンプルである。この情報とはDNAにコードされた遺伝情報の他、RNAにコードされた遺伝情報を含む。混合核酸サンプルとしては、cfDNA、cfRNAを含むサンプルが挙げられ、具体的には、全血、血漿、血清、尿が挙げられ、より好ましくは全血、血漿、血清が挙げられる。

【0065】

混合核酸サンプルには、主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸とが含まれる。また、混合核酸サンプルにおける主要核酸と副次核酸の存在比率は、主要寄与体や副次寄与体のステータスによって変動し得る。

【0066】

ここでいう「主要寄与体」とは、出生前遺伝学的検査の場合には母親、癌検査の場合には検査対象者、移植臓器のモニタリングにおいてはレシピエントが該当する。つまり、「主要寄与体」とは混合核酸サンプルの取得由来である個体のことをいう。

また、「主要核酸」とは前記主要寄与体に関する遺伝情報を含む核酸である。主要核酸には、出生前遺伝学的検査の場合には母親ゲノムDNA若しくはその断片又は母親ゲノムDNAからの転写物であるRNA（母親由来のcfDNA又はcfRNA）、癌検査の場合には検査対象者のゲノムDNA若しくはその断片又は検査対象者のゲノムDNAからの転写物であるRNA（検査対象者由来のcfDNA又はcfRNA）、移植臓器のモニタリングにおいてはレシピエントのゲノムDNA若しくはその断片又はレシピエントのゲノムDNAからの転写物であるRNA（レシピエント由来のcfDNA又はcfRNA）が該当する。

【0067】

また「副次寄与体」とは、出生前遺伝学的検査の場合には胎児、癌検査の場合には癌細

胞、移植臓器のモニタリングにおいては移植臓器が該当する。つまり、「副次寄与体」とは主要寄与体の体内に存在する、主要寄与体の本来の遺伝情報とは異なる遺伝情報を有する個体、組織又は細胞のことをいう。

また、「副次核酸」とは前記副次寄与体に関する遺伝情報を含む核酸である。副次核酸には、出生前遺伝学的検査の場合には胎児ゲノムDNA若しくはその断片又は胎児ゲノムDNAからの転写物であるRNA（胎児由来のcfDNA又はcfRNA）、癌検査の場合には癌細胞のゲノムDNA若しくはその断片又は癌細胞のゲノムDNAからの転写物であるRNA（癌細胞由来のcfDNA又はcfRNA）、移植臓器のモニタリングにおいては移植臓器のゲノムDNA若しくはその断片又はドナーのゲノムDNAからの転写物であるRNA（移植臓器由来のcfDNA又はcfRNA）が該当する。

【0068】

なお、本発明のモデル関数の作成方法は、あくまでもモデル関数を作成することが目的であるため、主要核酸及び副次核酸を含む混合核酸サンプルは人工的なものであっても構わない。例えば、主要核酸を含む血液に、副次核酸を模した核酸をスパイク（添加）することにより、混合核酸サンプルを調製してもよい。

【0069】

工程A-1で用意するデータセットには、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットが含まれる。ここで、データセットに含まれる多型座位の数は特に限定されず、好ましくは5以上、より好ましくは10以上、さらに好ましくは15以上、さらに好ましくは18以上である。

【0070】

このデータセットは、多型座位における各アレルを区別して検出できる分析手段によって得られたものであれば特に限定されない。当該分析手段としては、好ましくは多型座位における一塩基置換（SNPs）を区別して検出できる分析手段が挙げられる。

分析手段としては、例えば、SNPsの検出に使用される塩基配列解析、デジタルPCR、マイクロアレイ、リアルタイムPCRなどが挙げられる。

【0071】

塩基配列解析の具体的な手段としては次世代シーケンサー（NGS）が挙げられる。次世代シーケンサーは、クローン的に増幅された分子及び単一核酸分子の大量の並列配列決定を可能にする配列決定方法である。本発明においては、何れのNGSシステムを採用しても構わない。例えばパイロシーケンシング（GS Junior（Roche社）など）、可逆的色素ターミネーターを使用する合成によるシーケンシング（MiSeq（Illumina社）など）、ライゲーションによるシーケンシング（SeqStudio Genetic Analyzer（Thermo Fisher SCIENTIFIC社）など）、イオン半導体シーケンシング（Ion Proton System（Thermo Fisher SCIENTIFIC社）など）、CMOS（相補型金属酸化膜半導体）チップによるシーケンシング（iSeq 100 System（Illumina社）など）などが挙げられる。

次世代シーケンサーによって読み込んだ配列データを解析し、多型座位における特定の配列（特定のSNPs）を有するアレルのリード数を、当該アレルの存在を示す信号として解釈することができる。

また、次世代シーケンサーに供するライブラリーの調製段階において、核酸分子を個別に識別可能にするバーコード配列（Unique Molecular Identifiers（UMI）、Unique Molecular Tag（UMT））を解析対象の核酸断片に連結させた場合、多型座位における特定の配列（特定のSNPs）を有するアレルであることを特定するUMTのカウント数を当該アレルの存在を示す信号として解釈することができる。

本発明における分析手段として次世代シーケンサーを採用する場合、あらかじめ既知である多型座位を特異的に増幅するターゲットシーケンス法を採用することが好ましい。

【0072】

デジタルPCRは、1ウェルあたりに核酸分子が1分子入るか、入らないかという程度となるように多数のウェルへサンプルを分配して個別にPCRを行う方法である。ターゲット配列を含むウェルではPCR増幅が進んで蛍光シグナルが検出されるが、ターゲット配列を含まないウェルではPCR増幅が進まず、蛍光シグナルは検出されない。PCR後、各ウェルでシグナル増幅の「ある(+) / なし(-)」を判別し、シグナルの「ある(+)」ウェル数をターゲットのコピー数として算出する。

デジタルPCRにSNPsなどの変異を精度よく判別可能なプローブ(TaqMan R プローブやサイクリングプローブなど)を組み合わせれば、特定の配列(特定のSNPs)を有するアレルが増幅されたウェルのみで蛍光が観察される。アレルごとに異なる発光波長を有する蛍光標識プローブを設計すれば、一つの多型座位に存在する異なるアレルを蛍光色によってそれぞれ区別して検出することができる。特定のアレルに対応する蛍光シグナルの「ある(+)」ウェル数を当該アレルの存在を示す信号として解釈することができる。

10

【0073】

マイクロアレイは、既知の配列を有するDNA、DNA断片、cDNA、オリゴヌクレオチド、RNAまたはRNA断片などの核酸をプローブとして、数百個~数十万個まで配列して固相化させ、プローブに相補的な配列を有する核酸がハイブリダイズした際に、これを蛍光標識により検出する方法である。SNPsタイピングを行うマイクロアレイを特にSNPアレイともいう。

一つの座位に複数のアレルが想定される場合、各アレルを別個に固相化することで、これらを区別して検出することが可能となる。特定のアレルが固相化されたポイントにおける蛍光強度を当該アレルの存在を示す信号として解釈することができる。

20

【0074】

リアルタイムPCRは、PCRによる核酸の増幅量に応じて生じる蛍光を分光蛍光光度計によりリアルタイムでモニターし解析する方法である。リアルタイムPCRにSNPsなどの変異を精度よく判別可能なプローブ(TaqMan R プローブやサイクリングプローブなど)を組み合わせたことが好ましい。アレルごとに異なる発光波長を有する蛍光標識プローブを設計すれば、一つの多型座位に存在する異なるアレルを蛍光色によってそれぞれ区別して検出することができる。

リアルタイムPCRによりデータセットを得ようとする場合、測定効率を向上させる観点からマルチプレックスPCRを採用することが好ましい。マルチプレックスPCRは、複数組のプライマーを使用し、複数のターゲット配列を一つの反応系中で一度に増幅する方法である。

30

リアルタイムPCRにおいては、特定のアレルに対応する蛍光シグナルの強度を当該アレルの存在を示す信号として解釈することができる。

【0075】

質量分析は、分子をイオン化し、その質量荷電比(m/z)を測定することによってイオンや分子の質量を測定する分析法である。本来は分子の質量を測定する方法であるが、特定の条件(特定のプライマーを使用してPCRを実施した場合や、特定の制限酵素で核酸分子を切断した場合など)で調製した核酸分子の質量が計測できれば、その質量をデータベースと照合することで、検出された核酸分子の塩基配列を同定することができる。このことから、質量分析はジェノタイピングに広く応用されている。

40

質量分析においては、特定のアレルを含む塩基配列に特有の m/z におけるイオン強度を当該アレルの存在を示す信号として解釈することができる。

【0076】

工程A-1で用意するデータセットは、上述したアレルの存在を示す信号の真偽が既知であることを要する。つまり、ある特定のアレルの存在を示す信号が検出された場合に、当該アレルの塩基配列を含む主要核酸又は副次核酸が混合核酸サンプルに含まれているかが既知であることを要する。

【0077】

50

なお、工程 A-1 はデータセットを用意する工程である。したがって、データセットを一次的に取得するための核酸分析の工程は、本発明の必須要素でない。言い換えると、工程 A-1 の具体的な実施態様として、本発明の実施者自身が核酸分析によって一次的にデータを取得することで上記データセットを用意する形態も当然に含まれるが、これに限られない。工程 A-1 の具体的な実施態様には、本発明の実施者以外の者が核酸分析によって一次的に取得したデータセットを二次的に取得することで上記データセットを用意する形態も当然に含まれる。

【0078】

[工程 A-2]

工程 A-2 は、上述のデータセットに含まれるデータについて主成分分析を行う工程である。具体的には、前記データセットに含まれるデータのうち、複数の多型座位の中で、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、以下の (A1) 及び (A2) を含む数値群を線形結合させ、1 つ以上の合成変数を生成する。

【0079】

(A1) は副次成分信号強度である。副次成分信号強度とは、副次核酸に由来する、特定の多型座位のアレルの存在を示す信号の強度である。

混合核酸サンプルの分析で互いに区別されて検出された 2 種類のアレルの存在を示す信号が、それぞれ主要核酸と副次核酸のどちらに由来するのか確認することは容易である。

循環無細胞核酸サンプルにおいては、ほとんどの場合、副次核酸よりも主要核酸の方が多く含まれるため、必然的に副次成分信号強度は、上述の主要成分信号強度よりも弱くなる。このような場合にはその内の信号強度が弱い方を副次成分信号強度とみなすことができる。

一方、例えば妊娠後期における母親由来核酸と胎児由来核酸の比率、また、癌が進行している場合の患者由来核酸と癌由来核酸の比率は、通常の場合と逆転していることがある。つまり、循環無細胞核酸サンプルにおける副次核酸の量が主要核酸の量と同等又はそれ以上となる場合があり得る。このような特殊な場合には、ジェノタイピングによって主要寄与体の遺伝型を予め特定しておき、混合核酸サンプルの分析結果と比較すればよい。これによって、混合核酸サンプルの分析によって互いに区別されて検出された 2 種類のアレルの存在を示す信号が、それぞれ主要核酸と副次核酸のどちらに由来するのか判断できる。

【0080】

データセットを一次的に取得した際に使用された測定手段の種類やその後のソフトウェアなどによるデータ処理によって、副次成分信号強度を表す単位や表現などは多岐に渡る。本発明においては副次核酸に由来する特定の多型座位のアレルの存在を示す信号強度を反映する数値の全てを包含して「副次成分信号強度」と呼ぶ。つまり、当該信号強度を直接的に表す数値の他、当該数値に定数を乗じた数値や、当該数値の累乗値及び累乗根など当該信号強度を反映する全ての数値が「副次成分信号強度」に含まれる。

【0081】

例えば、副次成分信号強度の元データを標準化した数値も「副次成分信号強度」との文言に含まれる。標準化の詳細は後述する。

【0082】

また、検出された他のパラメータに基づき副次成分信号強度の元データを加工して得られた数値も「副次成分信号強度」という文言に含まれる。副次成分信号強度の元データの加工のために用いられる「他のパラメータ」としてはノイズが挙げられる。ノイズの定義は後述の通りである。

【0083】

例えば、副次成分信号強度の元データから、解析の対象である複数の多型座位におけるノイズ強度又はその平均値を引いた数値も、「副次成分信号強度」として扱うことができる。ノイズ強度の平均値を求める際の母数は、ノイズが検出された多型座位の数であっても、解析対象とした全ての多型座位の数であってもよい。

【0084】

具体的には、ノイズが検出された多型座位及びノイズが検出されなかった多型座位の区別なく、一律に副次成分信号強度の元データから前記ノイズ強度の平均値を差し引く実施形態としてもよい。

また、ノイズが検出された特定の多型座位についてのみ、副次成分信号強度の元データから前記ノイズ強度の平均値を差し引く実施形態としてもよい。

また、ノイズが検出された特定の多型座位の副次成分信号強度から、当該特定の多型座位について検出されたノイズ強度を差し引く実施形態としてもよい。

【0085】

また、前記特定の多型座位のアレルの存在を示す副次成分信号強度を、前記複数の多型座位におけるノイズ強度の平均値により割った数値を「副次成分信号強度」として扱う実施形態としてもよい。

つまり以下の式で表される数値を「副次成分信号強度」として扱う実施形態としてもよい。

$$(\text{副次成分信号強度}) / (\text{ノイズ強度の平均値})$$

【0086】

この通り「副次成分信号強度」というとき、これは1種類の数値のみを表すものではなく複数種類の数値を包含している。したがって、工程A-2において線形結合の対象となる数値群には、1種類の「副次成分信号強度」のみが含まれている形態であってもよいし、2種以上の「副次成分信号強度」が含まれていてもよい。

【0087】

(A2)は副次成分混入率である。副次成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合である。つまり、「副次成分混入率＝副次成分信号強度／信号強度総和」という数式で表すことができる。

【0088】

上記(A1)における説明と同様の理由から、副次成分混入率を表す単位や表現などは多岐に渡る。本発明においては特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合を反映する数値の全てを包含して「副次成分混入率」と呼ぶ。つまり、当該割合を直接的に表す数値の他、当該数値に定数を乗じた数値や、当該数値の累乗値及び累乗根など当該割合を反映する全ての数値が「副次成分混入率」に含まれる。

【0089】

例えば、副次成分混入率の元データを標準化した数値も「副次成分混入率」との文言に含まれる。標準化の詳細は後述する。

【0090】

また、検出された他のパラメータに基づき副次成分混入率の元データを加工して得られた数値も「副次成分混入率」という文言に含まれる。副次成分混入率の元データの加工のために用いられる「他のパラメータ」としてはノイズが挙げられる。ノイズの定義は後述の通りである。

【0091】

例えば、副次成分混入率の元データから、解析の対象である複数の多型座位における信号強度総和に対するノイズ強度の割合(ノイズ混入率)又はその平均値を引いた数値も、「副次成分混入率」として扱うことができる。ノイズ混入率の平均値を求める際の母数は、ノイズが検出された多型座位の数であっても、解析対象とした全ての多型座位の数であってもよい。

【0092】

具体的には、ノイズが検出された多型座位及びノイズが検出されなかった多型座位の区別なく、一律に副次成分混入率の元データから前記ノイズ混入率の平均値を差し引く実施形態としてもよい。

また、ノイズが検出された特定の多型座位についてのみ、副次成分混入率の元データから前記ノイズ混入率の平均値を差し引く実施形態としてもよい。

また、ノイズが検出された特定の多型座位の副次成分混入率から、当該特定の多型座位について検出されたノイズ強度のノイズ混入率を個別に差し引く実施形態としてもよい。

【0093】

また、前記特定の多型座位の副次成分混入率を、複数の多型座位におけるノイズ強度の平均値により割った数値を「副次成分混入率」として扱う実施形態としてもよい。

つまり以下の式で表される数値を「副次成分混入率」として扱う実施形態としてもよい。

(副次成分混入率) / (ノイズ強度の平均値)

【0094】

この通り「副次成分混入率」というとき、これは1種類の数値のみを表すものではなく複数種類の数値を包含している。したがって、工程A-2において線形結合の対象となる数値群には、1種類の「副次成分混入率」のみが含まれている形態であってもよいし、2種類以上の「副次成分混入率」が含まれていてもよい。

【0095】

工程A-2において線形結合の対象となる数値群には、上述の(A1)及び(A2)以外の数値が含まれていてもよい。つまり、特定の多型座位に関する(A1)及び(A2)の他、当該特定の多型座位に関する種々の測定値ないし算出値を含む数値群に対して線形結合を行う。

以下に線形結合の対象となる数値群に含めてもよい数値(A3)～(A5)について説明を加える。なお、以下に掲げる(A3)～(A5)から選ばれる1種のみを前記数値群に含めてもよいし、任意に選択された2種類以上の数値を前記数値群に含めてもよい。また、(A3)～(A5)の全てを前記数値群に含めてもよい。

【0096】

(A3)は主要成分信号強度である。主要成分信号強度とは、主要核酸に由来する、特定の多型座位の1のアレルの存在を示す信号の強度である。

上述の通り循環無細胞核酸サンプルにおいては、ほとんどの場合、副次核酸よりも主要核酸の方が多く含まれるため、必然的に主要成分信号強度は、上述の副次成分信号強度よりも弱くなる。このような場合にはその内の信号強度が強い方を主要成分信号強度とみなすことができる。

一方、例えば妊娠後期における母親由来核酸と胎児由来核酸の比率、また、癌が進行している場合の患者由来核酸と癌由来核酸の比率は、通常の場合と逆転していることがある。つまり、循環無細胞核酸サンプルにおける副次核酸の量が主要核酸の量と同等又はそれ以上となる場合があり得る。このような特殊な場合には、ジェノタイピングによって主要寄与体の遺伝型を予め特定しておき、混合核酸サンプルの分析結果と比較すればよい。これによって、混合核酸サンプルの分析によって互いに区別されて検出された2種類のアレルの存在を示す信号が、それぞれ主要核酸と副次核酸のどちらに由来するのか判断できる。

【0097】

上記(A1)の説明と同様の理由により、当該信号強度を直接的に表す数値の他、当該数値に定数を乗じた数値や、当該数値の累乗値及び累乗根など当該信号強度を反映する全ての数値が「主要成分信号強度」に含まれる。工程A-2において線形結合の対象となる数値群には、1種類の「主要成分信号強度」のみが含まれている形態であってもよいし、2種類以上の「主要成分信号強度」が含まれていてもよい。

【0098】

(A4)は主要成分混入率である。主要成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する主要成分信号強度の割合である。つまり、「主要成分混入率 = 主要成分信号強度 / 信号強度総和」という数式で表すことができる。

上記(A2)の説明と同様の理由により、当該割合を直接的に表す数値の他、当該数値に定数を乗じた数値や、当該数値の累乗値及び累乗根など当該信号強度を反映する全ての数値が「主要成分混入率」に含まれる。工程A-2において線形結合の対象となる数値群には、1種類の「主要成分混入率」のみが含まれている形態であってもよいし、2種類以上の「主要成分混入率」が含まれていてもよい。

【0099】

(A5) はノイズである。

生物学的、遺伝学的または試験系上の理由から、ある特定の一塩基多型座位において塩基「A」または塩基「G」を示す信号は観察され得るが、塩基「T」及び塩基「C」を示す信号は観察し得ない状況を考える。このような状況としては、特定の一塩基多型座位において父親が「A」をホモ接合で保有し、母親が「G」をホモ接合で保有する場合における、胎児の前記特定の一塩基多型を観察する状況が例として挙げられる（当該特定の一塩基多型座位において胎児が「T」及び「C」を保有していることは通常考えられない）。このような状況において、本来は観察し得ない塩基である「T」及び「C」を示す信号が観察された場合、この信号を「ノイズ」であると本発明では定義する。

10

つまり、ノイズは、特定の多型座位のアレルに起因する信号強度総和から、前記主要成分信号強度及び前記副次成分信号強度を差し引いて求められるものであり、「信号強度総和－（主要成分信号強度＋副次成分信号強度）」の数式で表すことができる。

【0100】

上述の通り工程A－1で用意するデータセットは複数の多型座位に関するデータの集合である。そのため、いうまでもないが、工程A－1で用意するデータセットには、特定の多型座位に関する上記（A1）及び（A2）並びに他の数値データを一組とするデータが複数含まれることになる。

【0101】

なお、線形接合の対象となる数値群に含まれる数値データは標準化されていることが好ましい。標準化データは以下の式で求めることができる。

20

標準化データ＝〔（元データ）－（平均値）〕／（標本標準偏差）

【0102】

「主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位」とは、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号が混在していない多型座位をいう。

【0103】

例えば、出生前遺伝学的検査の場合、ある特定の多型座位において、母親がアレルAとアレルBをヘテロ接合として有している場合には、父親の遺伝型の如何に関わらず、cfDNAの解析の結果、母親のゲノムDNAに由来するアレルAとアレルBの信号が必ず検出されてしまう。アレルAとアレルBの信号の何れかには、胎児のcffDNAに起因する信号が混じっているはずであるが、これを母親のゲノムDNAに由来する信号と区別することができない。このようなデータを解析の基礎に加えるとモデル関数の精度が落ちてしまう。

30

【0104】

また、癌検査の場合、検査対象者が先天的に癌に関連する変異をホモ接合又はヘテロ接合として有していた場合、当該変異は必ずctDNAにも含まれるため、検査対象由来の信号と癌細胞由来の信号とが混じることになる。このようなデータを解析の基礎に加えるとモデル関数の精度が落ちてしまう。

40

【0105】

また、移植臓器の定着のモニタリングの場合、ある特定の多型座位において、レシピエントがアレルAとアレルBをヘテロ接合として有している場合には、ドナーの遺伝型の如何に関わらず、cfDNAの解析の結果、レシピエントのゲノムDNAに由来するアレルAとアレルBの信号が必ず検出されてしまう。アレルAとアレルBの信号の何れかには、ドナーのcffDNAに起因する信号が混じっているはずであるが、これをレシピエントのゲノムDNAに由来する信号と区別することができない。このようなデータを解析の基礎に加えるとモデル関数の精度が落ちてしまう。

【0106】

以上のことから工程A－2においては、データ解析の対象とする多型座位を「主要核酸

50

に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位」に限定している。工程 A-2 の解析対象とする多型座位は、副次核酸に由来するアレルの存在を示す信号に、主要核酸に由来するアレルの存在を示す信号が混じっている可能性が無い多型座位と言い換えてもよい。

【0107】

工程 A-2 では上述した数値群を対象として線形結合させて 1 つ以上の合成変数を生成する。線形結合の手段としては主成分分析が好ましく例示できる。なお、別手段で生成した合成変数であっても構わない。別手段で生成する合成変数であっても、これが主成分分析で生成し得る合成変数であることが好ましい。

【0108】

線形結合により生成された合成変数は以下の 1 次斉次多項式により表される。なお、 n はデータセットに含まれる数値のうち、線形結合の対象である数値群に含まれる数値の種類の数を表す 2 以上の整数である。 X_n は線形結合の対象である数値群に含まれる数値である。 a_{1n} は線形結合を行う数値を重みづける係数である。

$$Z_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1n} X_n$$

本発明の好ましい実施形態では、合成変数を表す 1 次斉次多項式において、副次成分信号強度又は副次成分混入率が最大に重みづけされる。

【0109】

生成できる合成変数の数は、線形結合の対象となる数値群に含まれる数値の種類が多いほど増加する。工程 A-2 で生成する合成変数の数は特に限定されない。

【0110】

少なくとも (A1) 及び (A2) を含む数値群の線形結合によって合成変数を生成する実施形態について説明したが、前記数値群の非線形結合によって合成変数を生成する実施形態としてもよい。ここで、非線形結合とは、各数値のべき乗、各数値同士の積、商及びこれらの数値を指数に持つ関数等を指す。

【0111】

工程 A-2 の線形結合により得られた合成変数は、信頼性値と相関する関係にある。この相関関係を利用してモデル関数を作成するのであるが、その具体的な工程として、本発明は以下の工程 A-3-1 と工程 A-4-1 を有する。

【0112】

〔工程 A-3-1〕

工程 A-3-1 は、線形結合により生成された合成変数に信頼性値を割り当てる工程である。

工程 A-3-1 に供する合成変数は特に限定されないが、線形結合の対象となった数値群を最もよく反映する合成変数が好ましく挙げられる。言い換えると線形結合の対象となった数値群に対して最も高い寄与率を示す合成変数が好適に例示できる。これは主成分分析における第 1 主成分に相当するものである。

【0113】

工程 A-3-1 においては、まず、線形結合により生成した合成変数を複数に区分する。つまり、合成変数をその数値の大きさに応じて複数に区分する。

区分の方法は特に制限されない。合成変数の大きさに応じて等間隔に区分をしてもよいが、各区分の全てに生成された合成変数が包含されるように区分することが好ましい。さらに好ましい形態では、合成変数の大きさに応じて直線的に区分をするのではなく、指数的に区分をすることが好ましい。なぜならば、生成した合成変数と確率を曲線回帰するとシグモイド曲線となるからである。

【0114】

区分の数は限定されないが、好ましくは 3 区分以上、より好ましくは 5 区分以上、さらに好ましくは 7 区分以上、さらに好ましくは 10 区分以上、さらに好ましくは 12 区分以上、さらに好ましくは 15 区分以上、さらに好ましくは 18 区分以上に分ける。

【0115】

次に、各区分に含まれる合成変数に対応する副次成分信号強度のうちこれが真であった割合を求める。つまり、各区分に含まれる全ての合成変数のうち、真であった副次成分信号強度に対応する合成変数の割合を求める。本明細書においては、この割合を「確率」と呼ぶ。

なお、副次成分信号強度は、副次核酸における当該多型座位に存在する特定のアレルの存在を示唆するものである。この副次成分信号強度による示唆通り、実際に副次核酸に当該特定のアレルが存在する場合には、これを「真」とする。

【0116】

合成変数の各区分における確率を求めた後、これを各区分に含まれる合成変数に対応する確率として付与する。具体的には、各区分を代表する一の合成変数の値に当該区分における確率を割り当てる。この工程により、合成変数と確率の散布図が作成可能な状態となる。

10

【0117】

[工程A-4-1]

工程A-4-1においては、上述した各区分に含まれる合成変数と、各区分に含まれる合成変数に対応する確率について回帰分析を行う。これにより合成変数を説明変数、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数を求める。

【0118】

「確率」と「信頼性値」は対応関係にある。本明細書においては、モデル関数を作成するために用いられるパラメータを「確率」と呼び、モデル関数に説明変数を入力することで算出されるパラメータを「信頼性値」と呼ぶ。

20

【0119】

工程A-4-1における回帰分析の手法は特に限定されないが、最小二乗法が好ましく例示できる。

モデル関数はシグモイド関数となる。合成変数が主成分分析における第1主成分である場合、モデル関数は以下の式1で表すことができる。

【0120】

【数1】

[式1]

30

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x0)}}$$

(x1: 第1主成分、 A1: 遷移領域の勾配、 x01: 半値点)

【0121】

なお、本発明においては、上記式1の場合に限らず、2個の媒介変数を有するシグモイド関数の形式で信頼性値を算出するためのモデル関数を作成することが好ましい。式1における媒介変数はA1とx01が該当する。

40

【0122】

式1中、A1は好ましくは15.4～15.6であり、より好ましくは15.5である。また、x01は好ましくは-0.8～-0.6であり、より好ましくは-0.9である。なお、小数点第2位を四捨五入したときに上記数値に該当するものは、ここで規定した数値範囲内に含まれるものとする。

【0123】

上述の方法により取得したモデル関数は極めて汎用性が高い。工程A-1で用意したデータセットの取得条件とは異なる条件で一次的に取得したデータセットの解析にも応用できる。例えば、工程A-1で用意したデータセットの取得条件との間で、サンプル量や濃

50

度の違い、解析した多型座位の違い、信号の種類（リード数やUMTカウント）の違いがある条件で一次的に取得したデータセットにおける信頼性値の算出に、当該モデル関数を応用できる。

つまり、別の条件で取得されたデータセットについて信頼性値の算出をしたい場合に、当該別の条件について改めてモデル関数を作成する必要が無い。いったん本発明の方法によりモデル関数を作成しておけば、別条件で取得されたデータセットの解析にも転用できる。

【0124】

さらには、モデル関数の作成の基礎としたデータセットとは異なる種類の検査によって得られたデータセットの解析にも応用できる。例えば、出生前遺伝学的検査に関するデータセットに基づいて作成したモデル関数は、癌検査や移植臓器の定着のモニタリングにおいて取得されたデータセットの解析に転用することができる。

【0125】

なお、モデル関数の作成のために用いた、線形結合に供した数値群に含まれる数値の種類及びその個数と、モデル関数への入力値とする合成変数を生成するための線形結合に供した数値群に含まれる数値の種類及びその個数は、一致していることが好ましい。

【0126】

以上、合成変数と信頼性値との相関関係に基づくモデル関数の作成方法について説明したが、本発明はこれに留まらず、別の指標を説明変数とする信頼性値を算出するためのモデル関数を提供することができる。本発明は、後述するモデル関数 $f_2(x_2)$ 、 $f_3(x_3)$ を作成する方法にも関する。以下、それぞれのモデル関数を作成する方法について詳述する。

【0127】

まずモデル関数 $f_2(x_2)$ を作成する方法について説明する。この方法は、工程A-1、工程A-3-2及び工程A-4-2を備える。工程A-1の内容については上述した通りである。以下、工程A-3-2及び工程A-4-2について説明する。

【0128】

[工程A-3-2]

工程A-3-2においては、まず上述の(A1)副次成分信号強度を複数に区分する。つまり、(A1)副次成分信号強度をその数値の大きさに応じて複数に区分する。

区分の方法は特に制限されない。副次成分信号強度の大きさに応じて等間隔に区分してもよいが、各区分の全てに副次成分信号強度が包含されるように区分することが好ましい。さらに好ましい形態では、副次成分信号強度の大きさに応じて直線的に区分をするのではなく、指数的に区分をすることが好ましい。なぜならば、副次成分信号強度と信頼性値を曲線回帰するとシグモイド曲線となるからである。

【0129】

区分の数は限定されないが、好ましくは3区分以上、より好ましくは5区分以上、さらに好ましくは7区分以上、さらに好ましくは10区分以上、さらに好ましくは12区分以上、さらに好ましくは15区分以上、さらに好ましくは18区分以上に分ける。

【0130】

次に、各区分に含まれる副次成分信号強度に対応する副次成分信号強度のうちこれが真であった割合を求める。つまり、各区分に含まれる全ての副次成分信号強度の数値のうち、真であった副次成分信号強度の割合を求める。本明細書においては、この割合を「確率」と呼ぶ。

なお、副次成分信号強度は、副次核酸における当該多型座位に存在する特定のアレルの存在を示唆するものである。この副次成分信号強度による示唆通り、実際に副次核酸に当該特定のアレルが存在する場合には、これを「真」とする。

【0131】

副次成分信号強度の各区分における確率を求めた後、これを各区分に含まれる副次成分信号強度に対応する確率として付与する。具体的には、各区分を代表する一の副次成分信

10

20

30

40

50

号強度の値に当該区分における確率を割り当てる。この工程により、副次成分信号強度と確率の散布図が作成可能な状態となる。

【0132】

[工程A-4-2]

工程A-4-2においては、上述した各区分に含まれる副次成分信号強度と、各区分に含まれる副次成分信号強度に対応する確率について回帰分析を行う。これにより副次成分信号強度を説明変数 x_2 、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数 $f_2(x_2)$ を求める。工程A-4-2における回帰分析の手法は特に限定されないが、最小二乗法が好ましく例示できる。

モデル関数 $f_2(x_2)$ はシグモイド関数となり、以下の式2で表すことができる。

【0133】

【数2】

[式2]

$$f_2(x_2) = \frac{1}{1 + e^{-A_2(x_2 - x_{02})}}$$

(x_2 : 副次成分信号強度、 A_2 : 遷移領域の勾配、 x_{02} : 半値点)

【0134】

上述の方法により取得したモデル関数 $f_2(x_2)$ は極めて汎用性が高く、いったん本発明の方法によりモデル関数 $f_2(x_2)$ を作成しておけば、別条件で取得されたデータセットの解析にも転用できる。また、モデル関数 $f_2(x_2)$ の作成の基礎としたデータセットとは異なる種類の検査によって得られたデータセットの解析にも応用できる。

【0135】

式2中、 A_2 は好ましくは1.8~2.0であり、より好ましくは1.9である。また、 x_{02} は好ましくは2.5~2.7であり、より好ましくは2.6である。なお、小数点第2位を四捨五入したときに上記数値に該当するものは、ここで規定した数値範囲内に含まれるものとする。

【0136】

次にモデル関数 $f_3(x_3)$ を作成する方法について説明する。この方法は以下の工程A-3-3及び工程A-4-3を備える。

【0137】

[工程A-3-3]

工程A-3-3においては、まず上述の(A_2)副次成分混入率を複数に区分する。つまり、(A_2)副次成分混入率をその数値の大きさに応じて複数に区分する。区分の方法は特に制限されない。副次成分混入率の大きさに応じて等間隔に区分をしてもよいが、各区分の全てに副次成分混入率が包含されるように区分することが好ましい。

さらに好ましい形態では、副次成分混入率の大きさに応じて直線的に区分をするのではなく、指数的に区分をすることが好ましい。なぜならば、副次成分混入率と確率を曲線回帰するとシグモイド曲線となるからである。

【0138】

区分の数は限定されないが、好ましくは3区分以上、より好ましくは5区分以上、さらに好ましくは7区分以上、さらに好ましくは10区分以上、さらに好ましくは12区分以上、さらに好ましくは15区分以上、さらに好ましくは18区分以上に分ける。

【0139】

次に、各区分に含まれる副次成分混入率に対応する副次成分信号強度のうちこれが真であった割合を求める。つまり、各区分に含まれる全ての副次成分混入率の数値のうち、真

であった副次成分混入率の割合を求める。本明細書においては、この割合を「確率」と呼ぶ。

なお、副次成分混入率はその算出根拠として副次成分信号強度を含むが、この副次核酸における当該多型座位に存在する特定のアレルの存在を示唆するものである。この副次成分混入率の算出根拠である副次核酸信号強度による示唆通り、実際に副次核酸に当該特定のアレルが存在する場合には、これを「真」とする。

【0140】

副次成分混入率の各区分における確率を求めた後、これを各区分に含まれる副次各成分混入率に対応する確率として付与する。具体的には、各区分を代表する一の副次成分混入率の値に当該区分における確率を割り当てる。この工程により、副次成分混入率と信頼性値の散布図が作成可能な状態となる。

10

【0141】

[工程A-4-3]

工程A-4-3においては、上述した各区分に含まれる副次成分混入率と、各区分に含まれる副次成分混入率に対応する確率について回帰分析を行う。これにより副次成分混入率を説明変数 x_3 、信頼性値を目的変数とする、信頼性値を算出するためのモデル関数 $f_3(x_3)$ を求める。工程A-4-3における回帰分析の手法は特に限定されないが、最小二乗法が好ましく例示できる。

モデル関数 $f_3(x_3)$ はシグモイド関数となり、以下の式3で表すことができる。

【0142】

20

【数3】

[式3]

$$f_3(x_3) = \frac{1}{1 + e^{-A_3(x_3 - x_{03})}}$$

(x_3 : 副次成分混入率、 A_3 : 遷移領域の勾配、 x_{03} : 半値点)

【0143】

30

式3中、 A_3 は好ましくは9.3～9.5であり、より好ましくは9.4である。また、 x_{03} は好ましくは0.5～0.7であり、より好ましくは0.6である。なお、小数点第2位を四捨五入したときに上記数値に該当するものは、ここで規定した数値範囲内に含まれるものとする。

【0144】

上述した複数のモデル関数はそれぞれ単独でデータセットに含まれる副次成分信号強度の信頼性を評価するのに有用である。しかし、作成した複数のモデル関数を互いに乗じることで、より有用なモデル関数を作成することができる。

【0145】

例えば、工程A-2において、2つ以上の合成変数を生成し、工程A-3-1において、その2つ以上の合成変数のそれぞれについて、信頼性値の付与を行う。次いで工程A-4-1において、2つ以上の合成変数のそれぞれを説明変数とする互いに独立した2以上のモデル関数を作成する。この2以上のモデル関数を互いに乗じることで、乗算で表されるモデル関数を作成する実施の形態としても構わない。

40

【0146】

また、以下の3つのモデル関数から選択される2以上のモデル関数を互いに乗じ、乗算で表されるモデル関数を作成してもよい。本発明の好ましい形態では、以下の3つのモデル関数の全てを互いに乗じて、乗算により表されるモデル関数を作成してもよい。

・工程A-1、工程A-2、工程A-3-1及び工程A-4-1によって作成したモデル関数

50

- ・工程 A-1、工程 A-3-2 及び工程 A-4-2 によって作成したモデル関数
- ・工程 A-1、工程 A-3-3 及び工程 A-4-3 によって作成したモデル関数

【0147】

本発明の好ましい実施の形態では、以下の式 4 に示すように、上述したモデル関数 $f_1(x_1)$ 、モデル関数 $f_2(x_2)$ 、モデル関数 $f_3(x_3)$ を互いに乗じて作成したモデル関数を用いる。

【0148】

【数 4】

[式 4]

$$f(x_1, x_2, x_3) = f_1(x_1) * f_2(x_2) * f_3(x_3)$$

10

【0149】

<1-2> 出生前遺伝学的検査

次いで出生前遺伝学的検査より得られたデータセットからモデル関数の作成を行う実施形態について説明する。なお、「<1-1> 概要」の項目において説明した事項は本実施形態にも妥当する。そのため、説明が重複する部分については適宜省略しながら本実施形態について説明する。

【0150】

本実施形態においては、主要寄与体には母親、副次寄与体には母親の胎内にいる胎児、混合核酸サンプルには母親から採取された循環無細胞核酸サンプルが該当する。

「<1-1> 概要」の項目において説明した工程 A-1、工程 A-2、工程 A-3-1 及び工程 A-4-1 は、本実施形態における工程 A₁-1、工程 A₁-2、工程 A₁-3-1 及び工程 A₁-4-1 に相当する。以下、各工程について説明する。

【0151】

[工程 A₁-1]

工程 A₁-1 は、循環無細胞核酸サンプルの測定により得られるデータセットを用意する工程である。循環無細胞核酸サンプルには、母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸が含まれている。通常、循環無細胞核酸サンプルには、主要核酸が副次核酸よりも多く含まれる。一方、妊娠後期にはその含有比率が逆転することもある。

20

【0152】

このデータセットには、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号が含まれる。多型座位としては、ヒト個人識別 (Human Identification, HID) で用いられる一塩基多型 (SNPs) のある座位が好ましく挙げられる。HID で用いられる既知の SNPs はデータベース化されており、これらの SNPs のある多型座位を任意に選択することができる。

30

【0153】

なお、各アレルの存在を示す信号の真偽は既知であることが必要である。信号の真偽を既知の状態とする手段としては、出生後の子について確定的な遺伝子検査を行う方法が挙げられる。また、実父と母親について確定的な遺伝子検査を行った結果、ある多型座位において父母共にホモ接合として同型又は異型のアレルを有している場合には、当該多型座位における胎児のアレル型を確定的に特定することができる。

40

【0154】

[工程 A₁-2]

工程 A₁-2 は、データセットに含まれるデータのうち、複数の多型座位の中で、母親においてホモ接合であり、父親においてホモ接合であり、かつ、主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A₁) 及び (A₂) の数値について、線

50

形結合を行う工程である。

母親においてホモ接合であり、父親においてホモ接合である多型座位に絞っているため、主要成分信号強度と副次成分信号強度の両方に母親ゲノムDNA由来の信号が寄与することはあり得ない。

【0155】

〔工程A1-3-1〕

工程A1-3-1は、線形結合により生成した合成変数に信頼性値を割り当てる工程であり、上述した工程A-3-1の説明内容が全て妥当する。なお、副次成分信号強度の真偽については以下の通り判別する。

【0156】

母親においてホモ接合であり、父親（実父）においてホモ接合であり、かつ、母親と父親とで異型であるアレルについては、父親由来のアレルに起因する副次成分信号が、母親がホモ接合で有するアレルとは区別されて検出されるはずである。

したがって、前記アレルについて、主要成分信号と区別されて副次成分信号が検出された場合、当該副次成分信号を真とする。

また、前記アレルについて、主要成分信号と区別されて副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。これは、副次成分信号が非検出であったという結果が偽であるということである。

【0157】

一方、母親においてホモ接合であり、父親（実父）においてホモ接合であり、かつ、母親と父親とで同型であるアレルについては、父親由来のアレルは母親がホモ接合で有するアレルと区別して検出することはできない。

したがって、前記アレルについて主要成分信号と区別されて副次成分信号が検出された場合、当該副次成分信号を偽とする。

また、主要成分信号と区別されて前記副次成分信号が検出されなかった場合、当該副次成分信号を真とする。これは、副次成分信号が非検出であったという結果が真であるということである。

【0158】

〔工程A1-4-1〕

工程A1-4-1はモデル関数を求める工程であり、上述した工程A-4-1の説明内容が全て妥当する。

【0159】

また、当然ながら本実施形態においても、副次成分信号強度を説明変数 x_2 とするモデル関数 $f_2(x_2)$ 、副次成分混入率を説明変数 x_3 とするモデル関数 $f_3(x_3)$ を求めることが好ましい。本実施形態においてモデル関数を求める具体的態様は、上述した工程A-4-2及び工程A-4-3に関する説明が妥当する。

また、本実施形態においても、作成した複数のモデル関数を互いに乗じて、乗算で表されるモデル関数を作成してもよい。その具体的な実施形態は上述した通りである。

【0160】

<1-3>癌検査

次いで癌検査より得られたデータセットからモデル関数の作成を行う実施形態について説明する。なお、「<1-1>概要」の項目において説明した事項は本実施形態にも妥当する。そのため、説明が重複する部分については適宜省略しながら本実施形態について説明する。

【0161】

本実施形態においては、主要寄与体は癌に関連する変異が観察される多型座位において正常型のアレルを有する健常人、副次寄与体は癌細胞が該当する。

【0162】

なお、本実施形態において、混合核酸サンプルは、健常人に関する遺伝情報を含む主要核酸を含有する健常人より採取した核酸サンプルに、癌に関連する変異が導入された多型

10

20

30

40

50

座位の塩基配列情報を含む複数の核酸断片からなる副次核酸をスパイク（添加）することで人工的に調製したものである。

より具体的には、健常人から採取された循環無細胞核酸サンプルに、癌に関連する変異型のアレルの配列を含む核酸断片をスパイクして人工的に調製した混合核酸サンプルが好ましく挙げられる。

混合核酸サンプルは、健常人より採取した核酸サンプルに、人工的に合成した核酸断片をスパイクすることで調製してもよい。

また、健常人より採取した核酸サンプルに、癌細胞株若しくは癌組織、又はその核酸抽出物をスパイクすることで混合核酸サンプルを調製してもよい。

【0163】

混合核酸サンプルは、癌検査の検査対象者の循環無細胞核酸サンプルを模したものである。混合核酸サンプルにおける主要核酸と副次核酸の混合比率は特に制限されないが、混合核酸サンプルには、主要核酸が副次核酸よりも多く含まれるように調整することが好ましい。言い換えると、副次核酸における特定の座位に起因する信号が、主要核酸における当該座位に起因する信号よりも小さくなるように、副次核酸をスパイクすることが好ましい。

したがって、スパイクする副次核酸は主要核酸に対して、遺伝子のコピー数として好ましくは50%未満、より好ましくは40%以下、さらに好ましくは30%以下、さらに好ましくは20%以下、さらに好ましくは10%以下とする。

【0164】

スパイクする核酸断片は、癌に関連する変異が含まれていればその断片長は特に制限されないが、好ましくは50～500bp、より好ましくは100～300bp、さらに好ましくは120～200bpのものが好適に例示できる。

【0165】

癌に関連する変異、特に一塩基置換変異が観察される多型座位は多数知られており、これらはデータベース化されている。スパイクする核酸断片としては、これら既知の癌関連一塩基置換変異のうちから任意のものを複数選択することができる。

【0166】

「＜1-1＞概要」の項目において説明した工程A-1、工程A-2、工程A-3-1及び工程A-4-1は、本実施形態における工程A2-1、工程A2-2、工程A2-3-1及び工程A2-4-1に相当する。以下、各工程について説明する。

【0167】

〔工程A2-1〕

工程A2-1は、上述した副次核酸がスパイクされた混合核酸サンプルの測定により得られるデータを含むデータセットを用意する工程である。

なお、工程A2-1において用意するデータセットには、副次核酸がスパイクされておらず主要核酸のみを含む核酸サンプルの測定により得られるデータも含まれていてもよい。

【0168】

このデータセットには、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号が含まれる。多型座位としては、癌に関連することが知られている一塩基多型（SNPs）のある座位が好ましく挙げられる。癌に関連するSNPsはデータベース化されており、これらのSNPsのある多型座位を任意に選択することができる。

【0169】

〔工程A2-2〕

前記データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（A1）及び（A2）の数値について、線形結合を行う工程である。

【0170】

〔工程A2-3-1〕

10

20

30

40

50

工程 A 2-3-1 は、線形結合により得られた合成変数に信頼性値を割り当てる工程であり、上述した工程 A-3-1 の説明内容が全て妥当する。なお、副次成分信号強度の真偽については以下の通り判別する。

【0171】

前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルにスパイクした場合には、該核酸断片について副次成分信号が検出されるはずである。

したがって、この場合、該核酸断片について副次成分信号が検出された場合、当該副次成分信号は真とする。

また、該核酸断片について副次成分信号が検出されなかった場合には、当該副次成分信号は偽とする。これは、副次成分信号が非検出であったという結果が真であるということである。

10

【0172】

一方、前記変異が導入された前記多型座位の塩基配列情報を含む核酸断片を混合核酸サンプルに添加していない場合には、該核酸断片について副次成分信号は検出されないはずである。

したがって、この場合、該核酸断片について副次成分信号が検出されたとき、当該副次成分信号は偽とする。

また、この場合、該核酸断片について副次成分信号が検出されなかったとき、当該副次成分信号は真とする。これは、副次成分信号が非検出であったという結果が真であるということである。

20

【0173】

[工程 A 2-4-1]

工程 A 2-4-1 はモデル関数を求める工程であり、上述した工程 A-4-1 の説明内容が全て妥当する。

【0174】

また、当然ながら本実施形態においても、副次成分信号強度を説明変数 x_2 とするモデル関数 $f_2(x_2)$ 、副次成分混入率を説明変数 x_2 とするモデル関数 $f_2(x_2)$ を求めることが好ましい。本実施形態においてモデル関数を求める具体的な態様は、上述した工程 A-4-2 及び工程 A-4-3 に関する説明が妥当する。

また、本実施形態においても、作成した複数のモデル関数を互いに乗じて、乗算で表されるモデル関数を作成してもよい。その具体的な実施形態は上述した通りである。

30

【0175】

また、癌検査より得られたデータセットからモデル関数の作成を行う別の実施形態について説明する。本実施形態の特徴は、単一の多型座位に関するデータに基づきモデル関数の作成を行う点にある。

具体的には、以下の工程 A 2'-1、工程 A 2'-2 及び上述の工程 A 2-3-1 及び工程 A 2-4-1 を含む。以下詳述するが、上述した他の実施形態についての説明が妥当する事項については説明を省略する。

【0176】

[工程 A 2'-1]

工程 A 2'-1 は、上述した副次核酸が互いに異なる含有割合でスパイクされた複数の混合核酸サンプルの測定により得られるデータセットを用意する工程である。工程 A 2-1 との違いは、副次核酸が互いに異なる含有割合でスパイクされた複数の混合核酸サンプルを用意する点にある。

40

また、上述した工程 A 2-1 は複数の多型座位に関するデータが含まれる一方、工程 A 2'-1 のデータセットにおいては、主要核酸及び副次核酸における、単一の多型座位における各アレルの存在を示す信号が含まれていればよい点でも相違する。

つまり、工程 A 2'-1 は、単一の多型座位に関するデータを用意すればよい一方、副次核酸の含有割合が互いに異なる複数の混合核酸サンプルに関するデータを用意する点に特徴がある。

50

【0177】

〔工程A₂'-2〕

工程A₂'-2は、データセットに含まれるデータのうち、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された単一の多型座位に関する、少なくとも以下の(A₁')及び(A₂')を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程である。

(A₁') 前記副次核酸に由来する前記単一の多型座位のアレルの存在を示す副次成分信号強度。

(A₂') 前記単一の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率

10

【0178】

なお、(A₁')と(A₂')は、工程A₂'-1で用意するデータが単一の多型座位に関するデータであることから表現上の差異が生じているに過ぎず、その本質は上で説明した(A₁)と(A₂)と同一である。

【0179】

続く工程A₂-3-1及び工程A₂-4-1は上述した通りであるため、詳細な説明は省略する。

【0180】

なお、工程A₂'-1、工程A₂'-2及び上述の工程A₂-3-1及び工程A₂-4-1を含む実施形態は、検量線を作成する一般的な手法が無いマイクロアレイやデジタルPCR、塩基配列決定手段（特に次世代シーケンサー）により取得されたデータからモデル関数を作成する場合に有用である。

20

【0181】

<1-4> 移植臓器の定着のモニタリング

次いで移植臓器の定着のモニタリングより得られたデータセットからモデル関数の作成を行う実施形態について説明する。なお、「<1-1> 概要」の項目において説明した事項は本実施形態にも妥当する。そのため、説明が重複する部分については適宜省略しながら本実施形態について説明する。

【0182】

本実施形態においては、主要寄与体が臓器移植のレシピエント、副次寄与体がドナーから移植される移植臓器に該当する。

30

【0183】

本実施形態における混合核酸サンプルは、レシピエントに関する遺伝情報を含む主要核酸と、移植臓器に関する遺伝情報を含む副次核酸を含む。混合核酸サンプルにおいては、主要核酸が副次核酸よりも多く含まれる。なお当然のことながら移植臓器に関する遺伝情報はドナーに関する遺伝情報と一致する。

なお、混合核酸サンプルは、移植後のレシピエントから取得したサンプル、具体的には循環無細胞核酸サンプルであってもよい。

または、レシピエントから取得したレシピエント由来の主要核酸と、ドナー又は移植臓器から取得したドナー由来の副次核酸を人工的に混合することで調製してもよい。この場合、主要核酸に起因する信号が副次核酸に起因する信号よりも強く検出されるように、主要核酸に対して副次核酸をコピー数として好ましくは50%未満、より好ましくは40%以下、さらに好ましくは30%以下、さらに好ましくは20%以下、さらに好ましくは10%以下の割合で混合する。

40

【0184】

「<1-1> 概要」の項目において説明した工程A-1、工程A-2、工程A-3-1及び工程A-4-1は、本実施形態における工程A₃-1、工程A₃-2、工程A₃-3-1及び工程A₃-4-1に相当する。以下、各工程について説明する。

【0185】

〔工程A₃-1〕

50

工程 A 3-1 は、上述した混合核酸サンプルの測定により得られるデータセットを用意する工程である。

このデータセットには、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号が含まれる。多型座位としては、ヒト個人識別 (Human Identification, HID) で用いられる一塩基多型 (SNPs) のある座位が好ましく挙げられる。HID で用いられる既知の SNPs はデータベース化されており、これらの SNPs のある多型座位を任意に選択することができる。

【0186】

なお、各アレルの存在を示す信号の真偽は既知であることが必要である。信号の真偽を既知の状態とする手段としては、レシピエントとドナーそれぞれのゲノム DNA を解析することでそれぞれの遺伝型を特定する方法が挙げられる。これによりレシピエントとドナーの双方が有していないアレルの存在を示す信号が得られた場合には、これが偽であると判別できる。

【0187】

また、レシピエントから取得したレシピエント由来の主要核酸と、ドナー又は移植臓器から取得したドナー由来の副次核酸を人工的に混合することで調製した混合核酸サンプルについてのデータセットであれば、レシピエントが有しておらず、ドナーがヘテロ接合又はホモ接合として有しているアレルの存在を示す信号が得られたとき、これを真であると判別することができる。

【0188】

工程 A 3-1 で用意するデータセットに、レシピエントの遺伝情報を有する主要核酸のみを含む核酸サンプルに関するデータを含めることも有用である。当該核酸サンプルにはドナーに由来する副次核酸が含まれていないため、レシピエントが有していないがドナーが有しているアレルの存在を示す信号が得られたとき、これを偽であると判別することができる。

【0189】

〔工程 A 3-2〕

工程 A 3-2 は、データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (A 1) 及び (A 2) の数値について、線形結合を行う工程である。

具体的には、ある多型座位において、レシピエントが特定のアレルをホモ接合として有している場合には、当該特定のアレル以外の別のアレルの存在を示す副次成分信号強度には、レシピエントのアレルに起因する信号が混じることはあり得ない。この場合には、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号は区別して検出される。

【0190】

〔工程 A 3-3-1〕

工程 A 3-3-1 は線形結合により生成された合成変数に信頼性値を割り当てる工程であり、上述した工程 A-3-1 の説明内容が全て妥当する。なお、副次成分信号強度の真偽については以下の通り判別する。

【0191】

レシピエントが有していないアレルであって、かつ、ドナーがホモ接合若しくはヘテロ接合で有しているアレルについては、レシピエントが有するアレルと区別されてドナーが有しているアレルに由来する副次成分信号が検出されるはずである。

したがって、前記アレルについて、主要成分信号と区別されて副次成分信号が検出された場合、当該副次成分信号を真とする。

また、前記アレルについて、主要成分信号と区別されて副次成分信号が検出されなかった場合、当該副次成分信号を偽とする。これは、副次成分信号が非検出であったという結果が偽であるということである。

10

20

30

40

50

【0192】

一方、レシピエント及びドナーの何れもが有していないアレルについては、レシピエントが有するアレルと区別されて副次成分信号が検出されることは無い。

したがって、前記アレルについて、主要成分信号と区別されて副次成分信号が検出された場合、当該副次成分信号を偽とする。

また、前記アレルについて、主要成分信号と区別されて副次成分信号が検出されなかった場合、当該副次成分信号を真とする。これは、副次成分信号が非検出であったという結果が真であるということである。

【0193】

[工程A3-4-1]

工程A3-4-1はモデル関数を求める工程であり、上述した工程A-4-1の説明内容が全て妥当する。

【0194】

また、当然ながら本実施形態においても、副次成分信号強度を説明変数 x_2 とするモデル関数 $f_2(x_2)$ 、副次成分混入率を説明変数 x_3 とするモデル関数 $f_3(x_3)$ を求めることが好ましい。本実施形態においてモデル関数を求める具体的態様は、上述した工程A-4-2及び工程A-4-3に関する説明が妥当する。

また、本実施形態においても、作成した複数のモデル関数を互いに乗じて、乗算で表されるモデル関数を作成してもよい。その具体的な実施形態は上述した通りである。

【0195】

<2>信頼性の算出方法

本発明は信頼性の算出方法にも関する。以下、本発明の信頼性の算出方法の具体的な実施態様について説明する。なお、上述したモデル関数の作成方法の説明の内容のうち本発明の信頼性の算出方法の説明に妥当する部分については適宜省略する。

【0196】

本発明の信頼性の算出方法は、モデル関数にその説明変数を入力することで、信頼性値を算出する信頼性値の算出方法である。ここでいうモデル関数とは、上述の方法で求めたモデル関数、式1~3の何れかのモデル関数、又は式1~3で表されるモデル関数からなる群から選択される2以上のモデル関数を互いに乗じ、乗算で表されるモデル関数が挙げられる。

【0197】

モデル関数に入力すべき数値は、それぞれのモデル関数において説明変数とされているものである。具体的には、以下の工程B-1で用意するデータセットに含まれる以下の(B1)、(B2)及び以下の工程B-2で求めた合成変数から選ばれる1又は2以上の数値を説明変数としてモデル関数に入力する。

【0198】

本発明の信頼性の算出方法は以下の工程B-1を備える。また、モデル関数に入力すべき数値が合成変数である場合には、以下の工程B-2によって合成変数を生成する。

以下、工程B-1、工程B-2及び工程B-3-1を備える実施形態について説明する。

【0199】

[工程B-1]

工程B-1は、主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸とを含む混合核酸サンプルの測定により得られるデータセットを用意する工程である。当該混合核酸サンプルには、主要核酸が副次核酸よりも多く含まれる。そして、当該データセットには、主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号が含まれる。

【0200】

前記データセットの取得方法は特に限定されない。後述する分析手段を利用して一次的に取得しても構わないし、第三者が一次的に取得したものを二次的に取得しても構わない。

データセットは、多型座位における各アレルを区別して検出できる分析手段によって得

10

20

30

40

50

られたものであれば特に限定されない。当該分析手段としては、好ましくは多型座位における一塩基置換（SNPs）を区別して検出できる分析手段が挙げられる。

【0201】

分析手段としては、例えば、SNPsの検出に使用される次世代シーケンサーや、デジタルPCR、マイクロアレイ、マルチプレキシングPCR、質量分析などが挙げられる。これらの具体的な内容については「＜1＞モデル関数の作成方法」の項目で説明した通りである。

【0202】

混合核酸サンプルの種類も限定されない。例えば、出生前遺伝学的検査のために取得された妊婦の血液等から取得された循環無細胞核酸サンプル（cfDNA、cfRNA）や、癌検査のために取得された検査対象者の血液等から取得された循環無細胞核酸サンプル（cfDNA、cfRNA）や、移植臓器の定着のモニタリングのために取得されたレシピエントの血液等から取得された循環無細胞核酸サンプル（cfDNA、cfRNA）が好ましく挙げられる。

【0203】

なお、本発明の信頼性の算出方法におけるデータセットには、複数の多型座位における各アレルの存在を示す信号が含まれるが、この「複数の多型座位」は、モデル関数の作成の基礎として使用された「複数の多型座位」と同一である必要は無く、その重複度も限定は無い。

【0204】

当該重複度は、モデル関数の作成の基礎として使用された「複数の多型座位」を基準として、好ましくは80%以下であってもよいし、より好ましくは70%以下であってもよいし、さらに好ましくは60%以下であってもよいし、さらに好ましくは50%以下であってもよい。

【0205】

また、当該重複度は、モデル関数の作成の基礎として使用された「複数の多型座位」を基準として、0%であってもよいし、好ましくは10%以上であってもよいし、さらに好ましくは20%以上であってもよいし、さらに好ましくは30%以上であってもよいし、さらに好ましくは40%以上であってもよい。

【0206】

【工程B-2】

工程B-2は、前記データセットに含まれるデータのうち、複数の多型座位の中で、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、以下の（B1）及び（B2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程である。

【0207】

（B1）は副次成分信号強度である。副次成分信号強度とは、副次核酸に由来する、特定の多型座位のアレルの存在を示す信号の強度である。その定義や具体的態様については上記（A1）に関する説明がそのまま妥当する。

【0208】

（B2）は副次成分混入率である。副次成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合である。つまり、「副次成分混入率＝副次成分信号強度／信号強度総和」という数式で表すことができる。その定義や具体的態様については上記（A2）に関する説明がそのまま妥当する。

【0209】

工程B-2において線形結合の対象となる数値群には、上述の（B1）及び（B2）以外の数値が含まれていてもよい。つまり、特定の多型座位に関する（B1）及び（B2）の他、当該特定の多型座位に関する種々の測定値ないし算出値を含む数値群に対して線形結合を行う。

以下に線形結合の対象となる数値群に含めてもよい数値（B3）～（B5）について説

10

20

30

40

50

明を加える。なお、以下に掲げる（Ｂ３）～（Ｂ５）から選ばれる１種のみを前記数値群に含めてもよいし、任意に選択された２種以上の数値を前記数値群に含めてもよい。また、（Ｂ３）～（Ｂ５）の全てを前記数値群に含めてもよい。

【０２１０】

（Ｂ３）は主要成分信号強度である。主要成分信号強度とは、主要核酸に由来する、特定の多型座位の１のアレルの存在を示す信号の強度である。その定義や具体的態様については上記（Ａ３）に関する説明がそのまま妥当する。

【０２１１】

（Ｂ４）は主要成分混入率である。主要成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する主要成分信号強度の割合である。つまり、「主要成分混入率＝主要成分信号強度／信号強度総和」という数式で表すことができる。その定義や具体的態様については上記（Ａ１）に関する説明がそのまま妥当する。

【０２１２】

（Ｂ５）はノイズである。その定義や具体的態様については上記（Ａ１）に関する説明がそのまま妥当する。

【０２１３】

上述の通り工程Ｂ－１で用意するデータセットは複数の多型座位に関するデータの集合である。そのため、いうまでもないが、工程Ｂ－１で用意するデータセットには、特定の多型座位に関する上記（Ｂ１）及び（Ｂ２）並びに他の数値データを一組とするデータが複数含まれることになる。

【０２１４】

なお、線形接合の対象となる数値群に含まれる数値データは標準化されていることが好ましい。「標準化」とは周知の通り、複数あるデータの平均をゼロ、分散が１になるように変換することであり、「基準化」や「正規化」と呼ばれることもある。標準化データは以下の式で求めることができる。

$$\text{標準化データ} = [(\text{元データ}) - (\text{平均値})] / (\text{標本標準偏差})$$

【０２１５】

「主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位」とは、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号が混在していない多型座位をいう。

【０２１６】

例えば、出生前遺伝学的検査の場合、ある特定の多型座位において、母親がアレルＡとアレルＢをヘテロ接合として有している場合には、父親の遺伝型の如何に関わらず、cfDNAの解析の結果、母親のゲノムDNAに由来するアレルＡとアレルＢの信号が必ず検出されてしまう。アレルＡとアレルＢの信号の何れかには、胎児のcffDNAに起因する信号が混じっているはずであるが、これを母親のゲノムDNAに由来する信号と区別することができない。このようなデータは本発明の分析対象から除外する。

【０２１７】

また、癌検査の場合、検査対象者が先天的に癌に関連する変異をホモ接合又はヘテロ接合として有していた場合、当該変異は必ずctDNAにも含まれるため、検査対象由来の信号と癌細胞由来の信号とが混じることになる。このようなデータは本発明の分析対象から除外する。

【０２１８】

また、移植臓器の定着のモニタリングの場合、ある特定の多型座位において、レシピエントがアレルＡとアレルＢをヘテロ接合として有している場合には、ドナーの遺伝型の如何に関わらず、cfDNAの解析の結果、レシピエントのゲノムDNAに由来するアレルＡとアレルＢの信号が必ず検出されてしまう。アレルＡとアレルＢの信号の何れかには、ドナーのcffDNAに起因する信号が混じっているはずであるが、これをレシピエントのゲノムDNAに由来する信号と区別することができない。このようなデータは本発明の

10

20

30

40

50

分析対象から除外する。

【0219】

以上のことから工程B-2においては、データ解析の対象とする多型座位を「主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位」に限定している。工程B-2の解析対象とする多型座位は、副次核酸に由来するアレルの存在を示す信号に、主要核酸に由来するアレルの存在を示す信号が混じっている可能性が無い多型座位と言い換えてもよい。

【0220】

工程B-2では上述した数値群を対象として線形結合させて1つ以上の合成変数を生成する。線形結合の手段としては主成分分析が好ましく例示できる。なお、別手段で生成した合成変数であっても構わない。別手段で生成する合成変数であっても、これが主成分分析で生成し得る合成変数であることが好ましい。

10

【0221】

生成できる合成変数の数は、線形結合の対象となる数値群に含まれる数値の種類が多いほど増加する。工程B-2で生成する合成変数の数は特に限定されない。

【0222】

以上のようにして得た数値をモデル関数に入力することで信頼性値を算出する工程が、以下の工程B-3-1～工程B-3-4である。

【0223】

[工程B-3-1]

20

工程B-3-1は、工程B-2における線形結合により生成した合成変数を、その合成変数を説明変数、信頼性値を目的変数とする上述したモデル関数に入力し、信頼性値を算出する工程である。なお、モデル関数の作成のために用いた、線形結合に供した数値群に含まれる数値の種類及びその個数と、モデル関数への入力値とする合成変数を生成するための線形結合に供した数値群に含まれる数値の種類及びその個数は、一致していることが好ましい。

【0224】

また、本発明は、上述の工程B-1及び以下の工程B-3-2を含むことを特徴とする、信頼性値の算出方法にも関する。

【0225】

30

[工程B-3-2]

工程B-3-2は、前記(B1)の副次成分信号強度を、上述したモデル関数 $f_2(x_2)$ に入力し、信頼性値を算出する工程である。データセットに一次的に含まれる副次成分信号強度をモデル関数 $f_2(x_2)$ に入力するという操作で簡便にデータの信頼性値を算出することができる。

【0226】

また、本発明は、上述の工程B-1及び以下の工程B-3-3を含むことを特徴とする、信頼性値の算出方法にも関する。

[工程B-3-3]

40

工程B-3-3は、前記(B2)の前記副次成分混入率を、上述したモデル関数 $f_3(x_3)$ に入力し、信頼性値を算出する工程である。副次成分混入率をモデル関数 $f_3(x_3)$ に入力するという操作で簡便にデータの信頼性値を算出することができる。

【0227】

また、本発明は上述の工程B-1及び下記工程B-3'を備えることを特徴とする、信頼性値の算出方法にも関する。

[工程B-3']

工程B-3'は、以下の3種の数値から選ばれる変数を、その変数を説明変数とし、かつ、信頼性値を目的変数とする、乗算で表されるモデル関数に入力して、信頼性値を算出する工程である。

(i) 上記工程B-2で生成した合成変数。

50

(i i) 前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、前記副次核酸に由来する特定の多型座位のアレルの存在を示す副次成分信号強度。

(i i i) 前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率。

【 0 2 2 8 】

なお、ここでいう乗算で表されるモデル関数とは、上述した通り以下の3つのモデル関数から選択される2以上のモデル関数を互いに乗じることで、乗算で表されるモデル関数である。

・工程 A - 1、工程 A - 2、工程 A - 3 - 1 及び工程 A - 4 - 1 によって作成したモデル関数

・工程 A - 1、工程 A - 3 - 2 及び工程 A - 4 - 2 によって作成したモデル関数

・工程 A - 1、工程 A - 3 - 3 及び工程 A - 4 - 3 によって作成したモデル関数

【 0 2 2 9 】

本発明の好ましい実施形態では、上記 $f_1(x_1)$ 、 $f_2(x_2)$ 、 $f_3(x_3)$ のそれぞれの説明変数に相当する変数を式 4 で表されるモデル関数に入力し、信頼性値を算出する。

【 0 2 3 0 】

以下、本発明の信頼性値の算出方法の更に詳細な実施形態について説明する。具体的には、非侵襲的出生前親子鑑定、癌検査、移植臓器の定着のモニタリング、そして疾患リスクを評価する非侵襲的出生前検査の順に説明する。

【 0 2 3 1 】

< 2 - 1 > 非侵襲的出生前親子鑑定のために信頼性値を算出する方法

まず、非侵襲的出生前親子鑑定のために信頼性値を算出する方法について説明する。本項目においては上記 < 2 > に記載した事項が妥当するため、説明が重複する部分については適宜省略する。

【 0 2 3 2 】

本実施形態においては、主要寄与体が母親、副次寄与体が母親の胎内にいる胎児、混合核酸サンプルが母親から採取された循環無細胞核酸サンプルに相当する。

そして、上で説明した工程 B - 1、工程 B - 2 及び工程 B - 3 - 1 は、それぞれ以下に説明する工程 B₁ - 1、工程 B₁ - 2 及び工程 B₁ - 3 - 1 に相当する。

【 0 2 3 3 】

[工程 B₁ - 1]

工程 B₁ - 1 は、母親に関する遺伝情報を含む主要核酸と、胎児に関する遺伝情報を含む副次核酸と、を含む循環無細胞核酸サンプルの測定により得られるデータセットを用意する工程である。当該データセットは、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットである。

ここでいう前記複数の多型座位は、ヒト個人識別 (H I D) で用いられる多型座位であることが好ましい。

【 0 2 3 4 】

[工程 B₁ - 2]

工程 B₁ - 2 は、データセットに含まれるデータのうち、複数の多型座位の中で、母親においてホモ接合であり、かつ、主要核酸に由来するアレルの存在を示す信号と、副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記 (B 1) 及び前記 (B 2) を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程である。なお、擬父における前記多型座位の遺伝型は、ホモ接合であってもヘテロ接合であってもよい。

10

20

30

40

50

【0235】

[工程B1-3-1]

工程B1-3-1は、工程B1-2で生成した合成変数を、その合成変数を説明変数とするモデル関数に入力し、信頼性値を算出する工程である。

【0236】

<2-2>癌検査のために信頼性値を算出する方法

次に癌検査のために信頼性値を算出する方法について説明する。本項目においても上記<2>に記載した事項が妥当するため、説明が重複する部分については適宜省略する。

【0237】

本実施形態においては、主要寄与体は検査対象者、副次寄与体は癌細胞、混合核酸サンプルは検査対象者から採取された循環無細胞核酸サンプルに相当する。また、上記工程B-1、工程B-2及び工程B-3-1は、それぞれ下で説明する工程B2-1、工程B2-2及び工程B2-3-1に相当する。

10

【0238】

[工程B2-1]

工程B2-1は、検査対象者に関する遺伝情報を含む主要核酸を含み、癌細胞に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットであり、前記主要核酸及び前記副次核酸における、癌に関連する複数の多型座位における各アレルの存在を示す信号を含むデータセットを用意する工程である。

ここで「副次核酸を含み得る」とは、循環無細胞核酸サンプルに副次核酸が含まれる可能性を完全に否定できない状況のことをいう。

20

【0239】

[工程B2-2]

工程B2-2は、データセットに含まれるデータのうち、複数の多型座位の中で、正常型のアレルの存在を示す信号と、変異型のアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記(B1)及び前記(B2)を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程である。

正常型のアレルとは癌に罹患していない健康人において通常みられるアレルのことであり、変異型のアレルとは癌に関連があるとされる変異が導入されたアレルのことをいう。

【0240】

工程B2-2においては、前記データセットに含まれるデータから、前記複数の多型座位の中で、検査対象者において変異型のアレルをホモ接合又はヘテロ接合として有している多型座位に関するデータを除外することが好ましい。このように検査対象者において先天的に有している変異型アレルのある多型座位に関するデータを除くことによって、副次成分信号が、検査対象者自身に由来する主要成分信号と混ざって検出されたデータが除外される。これによって算出する信頼性値の精度が向上する。

30

【0241】

[工程B2-3-1]

工程B2-3-1は、工程B2-2で生成した合成変数を、その合成変数を説明変数とするモデル関数に入力し、信頼性値を算出する工程である。

40

【0242】

<2-3>移植臓器の定着のモニタリングのために信頼性値を算出する方法

次に移植臓器の定着のモニタリングのために信頼性値を算出する方法について説明する。本項目においても上記<2>に記載した事項が妥当するため、説明が重複する部分については適宜省略する。

【0243】

本実施形態においては、主要寄与体は臓器移植のレシピエント、副次寄与体は移植臓器、混合核酸サンプルはレシピエントから採取された循環無細胞核酸サンプルに相当する。

また、上述した工程B-1、工程B-2及び工程B-3-1は、それぞれ以下に説明する工程B3-1、工程B3-2及び工程B3-3-1に相当する。

50

【0244】

〔工程B3-1〕

工程B3-1は、レシピエントに関する遺伝情報を含む主要核酸を含み、移植臓器に関する遺伝情報を含む副次核酸を含み得る、循環無細胞核酸サンプルの測定により得られるデータセットを用意する工程である。当該データセットには、主要核酸及び副次核酸における、複数の多型座位における各アレルの存在を示す信号が含まれる。ここでいう複数の多型座位は、ヒト個人識別（HID）で用いられる多型座位であることが好ましい。

【0245】

〔工程B3-2〕

工程B3-2は、データセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B1）及び前記（B2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する工程である。

【0246】

〔工程B3-3-1〕

工程B3-3-1は、工程B3-2で生成した合成変数を、その合成変数を説明変数とするモデル関数に入力し、信頼性値を算出する工程である。

【0247】

<2-4>疾患リスクを評価する非侵襲的出生前検査のために信頼性値を算出する方法

次に疾患リスクを評価する非侵襲的出生前検査のために信頼性値を算出する方法について説明する。本項目においても上記<2>に記載した事項が妥当するため、説明が重複する部分については適宜省略する。

【0248】

本実施形態においては、主要寄与体が母親、副次寄与体が母親の胎内にいる胎児、混合核酸サンプルが母親から採取された循環無細胞核酸サンプルに相当する。

そして、上で説明した工程B-1、工程B-2及び工程B-3-1は、それぞれ以下に説明する工程B4-1、工程B4-2及び工程B4-3-1に相当する。

【0249】

〔工程B4-1〕

工程B4-1では、母親に関する遺伝情報を含む主要核酸と、母親の胎内にいる胎児に関する遺伝情報を含む副次核酸とを含む、前記母親から採取された循環無細胞核酸サンプルの測定により得られるデータセットを用意する。当該データセットには、主要核酸及び副次核酸における、疾患に関連する複数の多型座位における各アレルの存在を示す信号が含まれる。

【0250】

〔工程B4-2〕

工程B4-2では、まずデータセットに含まれるデータから、前記複数の多型座位の中で、母親において変異型のアリルをヘテロ接合として有している多型座位に関するデータを除外する。

【0251】

そして、除外後に残ったデータセットに含まれるデータのうち、前記複数の多型座位の中で、前記主要核酸に由来するアレルの存在を示す信号と、前記副次核酸に由来するアレルの存在を示す信号と、が区別して検出された多型座位に関する、少なくとも前記（B1）及び前記（B2）を含む数値群を線形結合させ、1つ以上の合成変数を生成する。

【0252】

〔工程B4-3-1〕

工程B4-3-1は、前記工程B-2で生成した前記合成変数を、その合成変数を説明変数とするモデル関数に入力し、信頼性値を算出する工程である。

【0253】

<3>除外条件の設定方法

10

20

30

40

50

上述した信頼性値の算出方法によれば、データセットに含まれる副次核酸における特定の多型座位における特定のアレルの存在を示す信号の信頼性 (F i d e l i t y) を評価することができる。

ただ、上述した信頼性値の算出方法は、副次核酸に由来する特定のアレルが混合核酸サンプルに含まれているにも関わらず、当該アレルの存在を示す信号の信頼性値が低く算出されてしまうケースがある。反対に副次核酸に由来する特定のアレルが混合核酸サンプルに含まれていないにも関わらず、当該アレルの存在を示す信号の信頼性値が高く算出されてしまうケースがある。このような例外結果は、分析対象となるデータセットに外れ値が含まれていることに起因する。これらの例外結果を除外することができれば、より高精度に信頼性値の算出が可能となる。

本発明の除外条件の設定方法は、モデル関数に入力すべき説明変数のデータを絞るため、データセットのうち除外すべきものを判定する除外条件を設定する方法に関する。本発明の除外条件の設定方法は、特に出生前遺伝学的検査に関するものである。

【0254】

具体的には、父母がそれぞれホモ接合で有している互いに異型の遺伝子座についての副次成分信号強度の信頼性値が、好ましくは0.8未満、より好ましくは0.9未満、さらに好ましくは0.99未満、さらに好ましくは0.999未満のものを除外するように、除外条件を設定することが好ましい。

また、父母がそれぞれホモ接合で有している互いに同型の遺伝子座についての副次成分信号強度の信頼性値が、好ましくは0.2以上、より好ましくは0.1以上、さらに好ましくは0.01以上、さらに好ましくは0.001以上のものを除外するように、除外条件を設定することが好ましい。

【0255】

以下、本発明の除外条件の設定方法について実施形態ごとに説明を加える。

【0256】

<3-1>除外条件の設定方法（実施形態1）

本発明の除外条件の設定方法の一つの実施形態は、以下の工程C-1-1、工程C-2-1、工程C-3-1及び工程C-4-1を備える。本実施形態により設定する除外条件は、上述した移植臓器の定着のモニタリングのために信頼性値を算出する方法に適用することができる。

【0257】

[工程C-1-1]

工程C-1-1は、主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの測定により得られるデータセットを用意する工程である。当該データセットには、前記主要核酸及び前記副次核酸における、複数の多型座位における各アレルの存在を示す信号を含むデータセットが含まれる。なお、前記信号の真偽は既知である。

【0258】

前記多型座位としては、ヒト個人識別 (H I D) で用いられる一塩基多型座位が好適に例示できる。

【0259】

なお、主要寄与体、副次寄与体、混合核酸サンプルは、以下の何れかに該当する。

(i) 前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルである。

(i i) 前記主要寄与体がレシピエントであり、前記副次寄与体が前記移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。

【0260】

[工程C-2-1]

工程C-2-1は、工程C-1-1で用意したデータセットのうち特定の条件に当ては

10

20

30

40

50

まる多型座位に関する数値を含む数値群を線形結合させて得られる合成変数のうち、最も寄与率が高い合成変数を生成する工程である。最も寄与率が高い合成変数とは、主成分分析を行う場合には第1主成分が該当する。

【0261】

工程C-2-1では、母親においてホモ接合であり、父親においてホモ接合であり、かつ、母親と父親とで異型であるアレル、または、レシピエントにおいてホモ接合であり、移植臓器のドナーにおいてホモ接合であり、かつ、レシピエントとドナーとで異型であるアレルの存在する多型座位に関する、少なくとも以下の(C1)、(C2)及び(C3)を含む数値群について線形結合を行う。

【0262】

(C1)は副次成分信号強度である。副次成分信号強度とは、副次核酸に由来する、特定の多型座位のアレルの存在を示す信号の強度である。その定義や具体的態様については上記(A1)に関する説明がそのまま妥当する。

【0263】

(C2)は副次成分混入率である。副次成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する副次成分信号強度の割合である。つまり、「副次成分混入率＝副次成分信号強度／信号強度総和」という数式で表すことができる。その定義や具体的態様については上記(A2)に関する説明がそのまま妥当する。

【0264】

(C3)はノイズである。ノイズは、特定の多型座位のアレルに起因する信号強度総和から、主要成分信号強度及び副次成分信号強度を差し引いて求められる数値である。その定義や具体的態様については上記(A5)に関する説明がそのまま妥当する。

【0265】

工程C-2-1において線形結合の対象となる数値群には、上述の(C1)、(C2)及び(C3)以外の数値が含まれていてもよい。つまり、特定の多型座位に関する(C1)、(C2)及び(C3)の他、当該特定の多型座位に関する種々の測定値ないし算出値を含む数値群に対して線形結合を行う。

以下に線形結合の対象となる数値群に含めてもよい数値(C4)～(C5)について説明を加える。なお、以下に掲げる(C4)～(C5)から選ばれる1種のみを前記数値群に含めてもよいし、任意に選択された2種以上の数値を前記数値群に含めてもよい。また、(C4)～(C5)の全てを前記数値群に含めてもよい。

【0266】

(C4)は主要成分信号強度である。主要成分信号強度とは、主要核酸に由来する、特定の多型座位の一のアレルの存在を示す信号の強度である。その定義や具体的態様については上記(A3)に関する説明がそのまま妥当する。

【0267】

(C5)は主要成分混入率である。主要成分混入率とは、特定の多型座位のアレルに起因する信号強度総和に対する主要成分信号強度の割合である。つまり、「主要成分混入率＝主要成分信号強度／信号強度総和」という数式で表すことができる。その定義や具体的態様については上記(A1)に関する説明がそのまま妥当する。

【0268】

ここで、データセットは複数の多型座位に関するデータの集合である。そのため、いうまでもないが、当該データセットには、特定の多型座位に関する上記(C1-1)～(C5-1)の数値データを一組とするデータが複数組含まれることになる。

なお、線形結合の対象となる数値群に含まれる数値データは標準化されていることが好ましい。

【0269】

なお、モデル関数の作成のために用いた、線形結合に供した数値群に含まれる数値の種類及びその個数と、工程C-2-1において合成変数を生成するための線形結合に供する数値群に含まれる数値の種類及びその個数は、一致していることが好ましい。

10

20

30

40

50

【0270】

[工程C-3-1]

工程C-3-1は、工程C-2-1における線形結合により得られた合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程である。その具体的な態様は特に限定されない。

【0271】

なお、前記外れ値は、本発明の方法で作成したモデル関数に入力して信頼性値を算出したときに異常な値を示す数値のことである。

具体的には、副次核酸に由来する特定のアレルが混合核酸サンプルに含まれているにも関わらず、当該アレルの存在を示す信号の信頼性値が、好ましくは0.6未満、より好ましくは0.7未満、さらに好ましくは0.8未満として算出されてしまう場合における当該アレルに関する数値を外れ値として扱うことができる。

10

また、副次核酸に由来する特定のアレルが混合核酸サンプルに含まれていないにも関わらず、当該アレルの存在を示す信号の信頼性値が、好ましくは0.4以上、より好ましくは0.3以上、さらに好ましくは0.2以上として算出されてしまう場合における当該アレルに関する数値を外れ値として扱うことができる。

【0272】

また、合成変数の平均値から、その標準偏差の好ましくは2倍以上、より好ましくは3倍以上、さらに好ましくは4倍以上、さらに好ましくは5倍以上の値をもって離れた数値を外れ値として扱うこともできる。

20

【0273】

工程C-3-1の具体的な態様として以下の方法が挙げられる。

まず、前記合成変数について仮の閾値を設定し、以下の仮の除外条件C1を設定する。

(仮の除外条件C1)

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも前記(C1)、前記(C2)及び前記(C3)を含む数値群を線形結合させて得られた、最も寄与率が高い合成変数が、仮の閾値未満に該当するデータセットは除去する。

30

【0274】

そして、解析対象であるデータセットにこの仮の除外条件C1を適用し、除外されずに残ったデータセットについて、上述の信頼性値の算出方法の発明を適用し、信頼性値を算出する。この算出された信頼性値の結果から例外結果が除外されているか否かを検定する。例外結果が除外されていない場合や、事実を正確に反映した信頼性値の結果が過剰に除外されている場合には、再び仮の除外条件を再設定し、上記と同様に検定を繰り返し、最適な条件を特定する。

40

【0275】

工程C-3-1は、後述する工程C-3-1-1及び工程C-3-1-2を含む形態としてもよい。

【0276】

[工程C-3-1-1]

工程C-3-1-1は、上述した本発明の方法により作成したモデル関数に、工程C-2-1における線形結合により生成した合成変数、(C1)副次成分信号強度、(C2)副次成分混入率及び(C3)ノイズのうち説明変数として必要な数値を入力して、信頼性値を算出する工程である。

50

信頼性値の算出に用いるモデル関数は、「＜１＞モデル関数の作成方法」の項目で説明したモデル関数であれば特に限定されない。好ましくは上述の式１～４の何れかで表されるモデル関数に説明変数を入力して、信頼性値を算出する。

【０２７７】

[工程Ｃ－３－１－２]

次いで工程Ｃ－３－１－２では、工程Ｃ－２－１における線形結合により生成した合成変数と、工程Ｃ－３－１－１で算出した信頼性値と、をプロットした散布図を作成する。例えば、縦軸に合成変数、横軸に信頼性値をプロットした散布図においては、横方向（信頼性値が広がる方向）に分散するデータポイントの集合（言い換えると合成変数の値の分散は小さく、信頼性値の値の分散が大きい集合）と、縦方向（合成変数が広がる方向）に分散するデータポイントの集合（言い換えると合成変数の値の分散が大きく、信頼性値の値の分散が小さい集合）が観察される。

10

このうち、信頼性値が広がる方向に分散するデータポイントの集合（横方向に延びる集合）を除外候補として特定する。

一方、合成変数が広がる方向に分散するデータポイントの集合（縦方向に延びる集合）を非除外候補として特定する。

そして、除外候補の一部又は全部を除外するように前記合成変数の値に閾値を設定する。

【０２７８】

除外されるデータポイントの割合が、除外候補の全データポイント（非除外候補と重複している部分も除外候補に含む）のうち、好ましくは５０％以上、より好ましくは６０％以上、さらに好ましくは７０％以上、さらに好ましくは８０％以上、さらに好ましくは９０％以上、さらに好ましくは９５％以上となるように、合成変数に閾値を設定する。

20

【０２７９】

[工程Ｃ－４－１]

工程Ｃ－４－１は、信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件Ｃ１として設定する工程である。

【０２８０】

（除外条件Ｃ１）

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの分析により得られたデータセットのうち、

30

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで異型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで異型であるアレル、の存在する多型座位に関する、少なくとも前記（Ｃ１）、前記（Ｃ２）及び前記（Ｃ３）を含む数値群を線形結合させて得られた、最も寄与率が高い合成変数が、前記工程Ｃ－３－１で設定した前記閾値未満に該当するデータセットは除去する。

【０２８１】

除外後に残ったデータセットに対して上述した本発明の信頼性値の算出方法を適用することで、例外結果が低減された高精度な信頼性値算出結果を得ることができる。

40

【０２８２】

＜３－２＞除外条件の設定方法（実施形態２）

本発明の除外条件の設定方法の一つの実施形態は、以下の工程Ｃ－１－２及び工程Ｃ－２－２、工程Ｃ－３－２及び工程Ｃ－４－２を備える。

【０２８３】

[工程Ｃ－１－２]

工程Ｃ－１－２は、主要寄与体に関する遺伝情報を含む主要核酸と、副次寄与体に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの測定により得られるデータセットを用意する工程である。当該データセットには、前記主要核酸及び前記副次核酸におけ

50

る、複数の多型座位における各アレルの存在を示す信号を含むデータセットが含まれる。
なお、前記信号の真偽は既知である。

【0284】

前記多型座位としては、ヒト個人識別（HID）で用いられる一塩基多型座位が好適に例示できる。

【0285】

なお、主要寄与体、副次寄与体、混合核酸サンプルは、以下の何れかに該当する。

（i）前記主要寄与体が母親であり、前記副次寄与体が前記母親の胎内にいる胎児であり、前記混合核酸サンプルが前記母親から採取された循環無細胞核酸サンプルである。

（ii）前記主要寄与体がレシピエントであり、前記副次寄与体が前記移植臓器であり、前記混合核酸サンプルが前記レシピエントから採取された循環無細胞核酸サンプルである。

【0286】

〔工程C-2-2〕

工程C-2-2は、工程C-1-2で用意したデータセットのうち特定の条件に当てはまる多型座位に関する数値を含む数値群を線形結合させて得られる合成変数のうち、一番目又は二番目に寄与率が高い合成変数を生成する工程である。一番目に寄与率が高い合成変数とは、主成分分析を行う場合には第1主成分が該当する。二番目に寄与率が高い合成変数とは、主成分分析を行う場合には第2主成分が該当する。

【0287】

工程C-2-2では、母親においてホモ接合であり、父親においてホモ接合であり、かつ、母親と前記父親とで同型であるアレル、または、レシピエントにおいてホモ接合であり、移植臓器のドナーにおいてホモ接合であり、かつ、レシピエントとドナーとで同型であるアレルの存在する多型座位に関する、少なくとも上述した（C1）、（C2）及び（C3）を含む数値群について線形結合を行う。なお、線形結合の対象となる数値群には（C1）、（C2）及び（C3）以外の数値が含まれていてもよく、例えば上述した（C4）～（C5）が挙げられる。その他、工程C-2-2の具体的態様については、上述した工程C-2-1における説明が妥当する。

【0288】

なお、モデル関数の作成のために用いた、線形結合に供した数値群に含まれる数値の種類及びその個数と、工程C-2-2において合成変数を生成するための線形結合に供する数値群に含まれる数値の種類及びその個数は、一致していることが好ましい。

【0289】

〔工程C-3-2〕

工程C-3-2は、工程C-2-2における線形結合により生成された合成変数の外れ値の一部又は全部を除外するように前記合成変数の値に閾値を設定する工程である。その具体的な態様は特に限定されない。外れ値の定義については、上述した工程C-3-1における説明が妥当する。

【0290】

工程C-3-2の具体的態様として具体的には以下の方法が挙げられる。

まず、前記合成変数について仮の閾値を設定し、以下の仮の除外条件C2を設定する。
（仮の除外条件C2）

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬父とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも前記（C1）、前記（C2）及び前記（C3）を含む数値群を線形

10

20

30

40

50

結合させて得られた、一番目又は二番目に寄与率が高い合成変数が、仮の閾値未満に該当するデータセットは除去する。

【0291】

そして、解析対象であるデータセットにこの仮の除外条件C2を適用し、除外されずに残ったデータセットについて、上述の信頼性値の算出方法の発明を適用し、信頼性値を算出する。この算出された信頼性値の結果から例外結果が除外されているか否かを検定する。例外結果が除外されていない場合や、事実を正確に反映した信頼性値の結果が過剰に除外されている場合には、再び仮の除外条件を再設定し、上記と同様に検定を繰り返し、最適な条件を特定する。

【0292】

工程C-3-2は、後述する工程C-3-2-1及び工程C-3-2-2を含む形態としてもよい。

【0293】

[工程C-3-2-1]

工程C-3-2-1は、上述した本発明の方法により作成したモデル関数に、工程C-2-2における線形結合により生成した合成変数、(C1)副次成分信号強度、(C2)の副次成分混入率及び(C3)ノイズのうち説明変数として必要な数値を入力して、信頼性値を算出する工程である。

信頼性値の算出に用いるモデル関数は、「<1>モデル関数の作成方法」の項目で説明したモデル関数であれば特に限定されない。好ましくは上述の式1~4の何れかで表されるモデル関数に説明変数を入力して、信頼性値を算出する。

【0294】

[工程C-3-2-2]

次いで工程C-3-2-2では、工程C-2-2における線形結合により生成した合成変数と、工程C-3-2-1で算出した信頼性値と、をプロットした散布図を作成する。例えば、縦軸に合成変数、横軸に信頼性値をプロットした散布図においては、横方向（信頼性値が広がる方向）に分散するデータポイントの集合（言い換えると合成変数の値の分散は小さく、信頼性値の値の分散が大きい集合）と、縦方向（合成変数が広がる方向）に分散するデータポイントの集合（言い換えると合成変数の値の分散が大きく、信頼性値の値の分散が小さい集合）が観察される。

このうち、合成変数が広がる方向に分散するデータポイントの集合（縦方向に延びる集合）を除外候補として特定する。

一方、信頼性値が広がる方向に分散するデータポイントの集合（横方向に延びる集合）を非除外候補として特定する。

そして、除外候補の一部又は全部を除外するように合成変数の値に閾値を設定する。

【0295】

除外されるデータポイントの割合が、除外候補の全データポイント（非除外候補と重複している部分も除外候補に含む）のうち、好ましくは50%以上、より好ましくは60%以上、さらに好ましくは70%以上、さらに好ましくは80%以上、さらに好ましくは90%以上、さらに好ましくは95%以上となるように、合成変数に閾値を設定する。

【0296】

[工程C-4-2]

工程C-4-2は、信頼性を算出するためのモデル関数に入力するデータセットから除外すべき条件を以下の除外条件C2として設定する工程である。

【0297】

(除外条件C2)

母親又はレシピエントに関する遺伝情報を含む主要核酸と、胎児又は移植臓器に関する遺伝情報を含む副次核酸とを含む、混合核酸サンプルの分析により得られたデータセットのうち、

母親においてホモ接合であり、擬父においてホモ接合であり、かつ、前記母親と前記擬

10

20

30

40

50

父とで同型であるアレル、又は

前記レシピエントにおいてホモ接合であり、前記移植臓器のドナーにおいてホモ接合であり、かつ、前記レシピエントと前記ドナーとで同型であるアレル、の存在する多型座位に関する、少なくとも前記（Ｃ１）、前記（Ｃ２）及び前記（Ｃ３）を含む数値群を線形結合させて得られた、一番目又は二番目に寄与率が高い合成変数が、前記工程Ｃ－３－２で設定した前記閾値未満に該当するデータセットは除去する。

【０２９８】

除外後に残ったデータセットに対して上述した本発明の信頼性値の算出方法を適用することで、例外結果が低減された高精度な信頼性値算出結果を得ることができる。

【０２９９】

<４>除外条件を適用して信頼性値を算出する方法

本発明は、上述の除外条件の設定方法により設定した除外条件Ｃ１及び／又は除外条件Ｃ２を上記「<２－３>移植臓器の定着のモニタリングのために信頼性値を算出する方法」における工程Ｂ２－１で用意するデータセットに適用して、除外されずに残ったデータセットについて、それぞれの信頼性値の算出方法の発明を適用し、信頼性値を算出する方法にも関する。

【０３００】

適用する除外条件は、除外条件Ｃ１と除外条件Ｃ２の何れか一方でもよいし両方であってもよい。なお、工程Ｂ１－２または工程Ｂ３－２において線形結合の対象とする数値群に含まれる数値の種類が、好ましくは１０種類以上、より好ましくは２０種類以上、さらに好ましくは３０種類以上である場合には、除外条件Ｃ１のみを適用するだけで、非常に精度の高い信頼性値の算出が可能となる。

【０３０１】

本発明の実施の形態は、除外条件Ｃ１及び／又は除外条件Ｃ２を適用する他は上記「<２－３>移植臓器の定着のモニタリングのために信頼性値を算出する方法」の項目で説明した内容をそのまま適用することができる。

【０３０２】

データセットに除外条件を適用しているため、これら入力値からは外れ値が除外されている。そのため、モデル関数により算出される信頼性値として例外結果が出力されてしまう問題を大きく低減することができる。

【０３０３】

<５>プログラム

本発明は、上述したモデル関数の作成方法、信頼性値の算出方法、除外条件の設定方法から選ばれる１又は２以上の方法をコンピュータに実行させるためのプログラムにも関する。コンピュータ内のプロセッサがハードディスク装置等の内蔵記憶装置に格納された本発明のプログラムに従って動作することにより、上述したモデル関数の作成方法、信頼性値の算出方法、除外条件の設定方法から選ばれる１又は２以上の方法を実行するように構成することができる。

【０３０４】

<６>記憶媒体

本発明は上述のプログラムを記録した記憶媒体にも関する。また、本発明は、上述した方法で作成されたモデル関数が記録された記憶媒体にも関する。記憶媒体としては、半導体メモリ、ハードディスク、磁気記憶媒体、光記憶媒体等、コンピュータにより読み出しが可能な記憶媒体が制限なく挙げられる。

【０３０５】

<７>信頼性値算出システム

本発明は、上述のモデル関数が記録された記憶部と、上述の信頼性値の算出方法を実行する処理部と、を備える、信頼性値算出システムにも関する。以下、本発明の信頼性値算出システムの好ましい実施形態について説明する。

【０３０６】

10

20

30

40

50

処理部は分析装置によって取得された鑑定対象となるデータセットをデータ処理する構成である。処理部は、例えば、記憶部に格納されたプログラム（上述の信頼性値の算出方法を実行するプログラム）を読みだして実行することで、信頼性値の算出に必要なデータ処理を実現する演算装置（計算機と称されてもよい）であってもよい。処理部は、データ処理の実行主体としての側面を有する。処理部として、例えば、CPU（Central Processing Unit）、MPU（Micro Processing Unit）、DSP（Digital Signal Processor）やFPGA（Field Programmable Gate Array）などが挙げられる。なお、処理部は、二以上のコアを含むマルチコアプロセッサであってもよい。

【0307】

記憶部は、処理部で実行される各種データ処理に係るデータやプログラムを記憶保持するように構成される回路である。記憶部は、不揮発性記憶装置と揮発性記憶装置の両方あるいは一方を少なくとも含んで構成される。例えば、RAM（Random Access Memory）、ROM（Read Only Memory）、SSD（Solid State Drive）、HDD（Hard Disk Drive）などが挙げられる。記憶部は、主記憶装置及び補助記憶装置などの各種記憶装置を総称したものである。プログラムは、予め記憶部に格納されていてもよいし、通信回路を介して接続された装置（サーバなど）からダウンロードされて記憶部に格納されてもよい。

【0308】

本実施形態の信頼性値算出システムは、上記工程B-1で用意したデータセットを入力する入力部を備える。入力部に入力した前記データセットは前記処理部に供される。前記処理部は記憶部に記憶された、上述の信頼性値の算出方法を実行するためのプログラムを読み出し、当該プログラムに従って、同じく記憶部に記憶されたモデル関数に前記データセットに含まれる又は前記データセットから生成した説明変数を入力して信頼性値を算出する。

【0309】

また本発明の好ましい実施の形態では、前記記憶部に上述の除外方法の設定方法により作成された除外条件C1及び／又は除外条件C2が記録されている。さらに好ましくは本実施形態の信頼性値算出システムは、上記工程B-1で用意したデータセットを入力する入力部を備える。

本実施形態では、入力部に入力した前記データセットは前記処理部に供される。前記処理部は記憶部に記憶された上述の除外条件C1及び／又は除外条件C2を読み出し、当該条件をデータセットに適用して、信頼性値の算出に適さないデータを除外する。処理部は、上述の信頼性値の算出方法を実行するためのプログラムを読み出し、当該プログラムに従って、同じく記憶部に記憶されたモデル関数に除外条件適用後に残った前記データセットに含まれる又は前記データセットから生成される説明変数を入力して信頼性値を算出する。

【実施例】

【0310】

<試験例1>モデル関数の作成

母親の口腔粘膜サンプル（母親の遺伝情報のみを含む）、父親の口腔粘膜サンプル（父親の遺伝情報のみを含む）及び、母親plasmaサンプル（母親及び、胎児の微量遺伝子を含む）を分析した次世代シーケンサー（NGS）による遺伝子配列検査データを1組のデータセットとし、全200組のデータセットを用意した。NGSは既知の184個のSNPsのある多型座位についてターゲットシーケンスを行ったものである。データセットには、36,800個（200組×184個）のSNPsに関するデータが含まれている。

【0311】

用意したデータセットに含まれる母親全血の解析データのうち、母親及び父親ともにホモ接合である多型座位に関するものだけを抽出した。これにより、10,415個のSN

10

20

30

40

50

P sに関するデータセットに絞られた。この抽出したデータセットに含まれる以下の5因子について主成分分析を行った。

- (1) 主要成分信号強度絶対値 [胎児Count Major]
- (2) 副次成分信号強度絶対値 [胎児Count minor]
- (3) 主要成分の混入率 (= (1) / シグナル強度総和) [胎児Freq. Major]
- (4) 副次成分の混入率 (= (2) / シグナル強度総和) [胎児Freq. minor]
- (5) ノイズ (= シグナル強度総和 - [(1) + (2)]) [胎児error]

【0312】

なお、上記(1)～(5)のデータを標準化した後に主成分分析を行った。

* (標準化データ) = [(生データ) - (平均値)] / (標本標準偏差)

【0313】

結果、第1主成分が信頼性値と高い相関を示す指標と成ることが明らかとなった。

【0314】

以下に記載する方法により各モデル関数を作成した。なお、モデル関数の作成には副次成分信号強度の真偽判定することが必要であるが、以下の規則に従って設定した正解に基づいて真偽判定を行った。

- ・ 父母の遺伝型がホモ且つ同型なら、胎児遺伝型はホモ (副次成分信号強度は偽)
- ・ 父母の遺伝型がホモ且つ異型なら、胎児遺伝型はヘテロ (副次成分信号強度は真)

【0315】

モデル関数 f 1 (x 1) の作成

主成分分析により得られた第1主成分をその大きさに応じて20に区分した。次に、各区分に含まれる第1主成分に対応する副次成分信号強度のうちこれが真であった割合 (確率) を求めた。そして各区分に含まれる第1主成分の代表値に、当該区分における確率を割り当てた。こうして得た第1主成分と信頼性値について最小二乗法を用いて回帰分析を行い、第1主成分を説明変数、信頼性値 (F i d e l i t y) を目的変数とするモデル関数 f 1 (x 1) を得た。回帰分析の寄与率 (R 2) は0.99以上と極めて良好であった。

図1にモデル関数 f 1 (x 1) を示すシグモイド曲線を示す。また、下の式5にモデル関数 f 1 (x 1) の式を示す。

【数5】

【式5】

$$f1(x1) = \frac{1}{1 + e^{-A1(x1-x01)}}$$

(x1: 第1主成分、 A1 = 15.5、 x 01 = -0.7)

【0316】

モデル関数 f 2 (x 2) の作成

副次成分信号強度絶対値をその大きさに応じて20に区分した。次に、各区分に含まれる副次成分信号強度絶対値のうちこれが真であった割合 (確率) を求めた。そして各区分における副次成分信号強度絶対値の代表値に、当該区分における確率を割り当てた。こうして得た副次成分信号強度絶対値と確率について最小二乗法を用いて回帰分析を行い、副次成分信号強度絶対値を説明変数、信頼性値 (F i d e l i t y) を目的変数とするモデル関数 f 2 (x 2) を得た。回帰分析の寄与率 (R 2) は0.99以上と極めて良好であった。

図2にモデル関数 f 2 (x 2) を示すシグモイド曲線を示す。また、下の式6にモデル関数 f 2 (x 2) の式を示す。

10

20

30

40

50

【数 6】

[式6]

$$f2(x2) = \frac{1}{1 + e^{-A2(x2-x02)}}$$

($x2$: 副次成分信号強度、 $A2 = 1.9$ 、 $x02 = 2.6$)

10

【0317】

モデル関数 $f3(x3)$ の作成

副次成分の混入率をその大きさに応じて20に区分した。次に、各区分に含まれる副次成分混入率に対応する副次成分信号強度のうちこれが真であった割合（確率）を求めた。そして各区分に含まれる副次成分混入率の代表値に、当該区分における確率を割り当てた。こうして得た副次成分混入率と確率について最小二乗法を用いて回帰分析を行い、副次成分混入率を説明変数、信頼性値（Fidelity）を目的変数とするモデル関数 $f3(x3)$ を得た。回帰分析の寄与率（ R^2 ）は0.99以上と極めて良好であった。

図3にモデル関数 $f3(x3)$ を示すシグモイド曲線を示す。また、下の式7にモデル関数 $f3(x3)$ の式を示す。

20

【数 7】

[式7]

$$f3(x3) = \frac{1}{1 + e^{-A3(x3-x03)}}$$

($x3$: 副次成分混入率、 $A3 = 9.4$ 、 $x03 = 0.6$)

30

【0318】

モデル関数 $f(x1, x2, x3)$ の作成

$f1(x1)$ 、 $f2(x2)$ 、 $f3(x3)$ を乗じて、以下の式4で表されるモデル関数 $f(x1, x2, x3)$ を作成した。

【0319】

【数 4】

[式4]

$$f(x1, x2, x3) = f1(x1) * f2(x2) * f3(x3)$$

40

【0320】

<試験例2> 信頼性値の算出

式4のモデル関数 $f(x1, x2, x3)$ を用いて、当該モデル関数の作成に使用した200セットのデータの信頼性を算出し結果の検証を行った。すなわち、混合核酸サンプルにおけるSNPsに係る座位についての第1主成分、副次成分信号強度絶対値、副次成分混入率をモデル関数 $f(x1, x2, x3)$ に入力し、その信頼性値を算出した。なお、信頼性値の算出においては、(1)と(2)の合計値が300未満のものは除外した8,148 SNPsに対して信頼性値（Fidelity）を計算した。

【0321】

50

算出した信頼性値 (F i d e l i t y) の分布図を図 4 示す。

左は父母がそれぞれホモ接合で有している互いに異型の S N P s に関する信頼性値を集計したものである (胎児遺伝型の正解はヘテロ接合)。

右は父母がそれぞれホモ接合で有している互いに同型の S N P s に関する信頼性値を集計したものである (胎児遺伝型の正解はホモ接合)。

【0322】

図 4 に示すように、本発明の方法によれば、精度よく S N P s に関する信号の信頼性を評価できる。

【0323】

<試験例 3> 除外条件の検討

図 4 左 (父母ホモ・異型) の集計においては、計算された信頼性値は 0.19 未満の値を示すものが分布し、一定数の例外事例が確認された。また、図 4 右 (父母ホモ・同型) の集計においては、0.9 以上の値を示すものが分布し、一定数の例外事例が確認された。

胎児遺伝型の正解がヘテロ接合であるにも関わらず算出された信頼性値が低いもの (例えば信頼性値 0.1 未満)、また胎児遺伝型の正解がホモ接合であるにも関わらず算出された信頼性値が高いもの (例えば信頼性値 0.9 以上) などの例外結果を除外するため、以下の方法で除外条件の詳細検討を行った。

【0324】

除外条件 1

父母がそれぞれホモ接合で有している互いに異型の S N P s (正解の胎児遺伝型はヘテロ接合) に関する混合核酸サンプルの分析により得られた S N P s (全 3, 196 個) について、上記 (1) ~ (5) のパラメータに対して主成分分析を行った。

一方で、主成分分析を行った (1) ~ (5) のパラメータに基づき上記モデル関数 $f(x_1, x_2, x_3)$ を利用して信頼性値を算出した。

次いで、主成分分析により得られた各主成分を y 軸、信頼性値を x 軸にプロットした散布図を作成した (図 5)。データの分散が最大であった第 1 主成分の散布図に示すように、x 軸方向に延びるデータポイントの集合と y 軸方向に延びるデータポイントの集合の 2 つが見られる。このうちの x 軸方向に延びるデータポイントの集合を除外すべき外れ値であるものとして第 1 主成分につき -1.9 の位置に閾値を設定した。この閾値に基づき以下の除外条件 1 を設定した。

(除外条件 1)

データセットのうち、父母がそれぞれホモ接合で有している互いに異型の S N P s についての上記 (1) ~ (5) の第 1 主成分が -1.9 未満のものは除外する。

【0325】

除外条件 2

父母がそれぞれホモ接合で有している互いに同型の S N P s に関する除外条件の適切な設定ができるか否か検討を行った。

父母がそれぞれホモ接合で有している互いに同型の S N P s (正解の胎児遺伝型はホモ接合) に関する混合核酸サンプルの分析により得られた S N P s (全 4, 952 個) について、上記 (1) ~ (5) のパラメータに対して主成分分析を行った。

一方で、主成分分析を行った (1) ~ (5) のパラメータに基づき上記モデル関数 $f(x_1, x_2, x_3)$ を利用して信頼性値を算出した。

次いで、主成分分析により得られた各主成分を y 軸、信頼性値を x 軸にプロットした散布図を作成した (図 6)。

データの分散が最大であった第 1 主成分の散布図に示すように、x 軸方向に延びるデータポイントの集合と y 軸方向に延びるデータポイントの集合の 2 つが見られる。このうちの y 軸方向に延びるデータポイントの集合を除外すべき外れ値であるものとして第 1 主成分につき 6 の位置に閾値を設定した。この閾値に基づき以下の除外条件 2 を設定した。

(除外条件 2)

データセットのうち、父母がそれぞれホモ接合で有している互いに同型の S N P s につい

10

20

30

40

50

ての上記(1)～(5)の第1主成分が6以上のものは除外する。

【0326】

＜試験例4＞信頼性値の再集計

200テストデータセットを用いた試験例3において設定した除外条件1、2に該当するSNPsに関するデータをデータセットから除外したうえで、試験例1と同手順で信頼性値を算出した(残ったSNPsの数: 8,081)。

算出した信頼性値(Fidelity)の分布図を図7に示す。

左は父母がそれぞれホモ接合で有している互いに異型のSNPsに関する信頼性値を集計したものである(胎児遺伝型の正解はヘテロ接合)。

右は父母がそれぞれホモ接合で有している互いに同型のSNPsに関する信頼性値を集計したものである(胎児遺伝型の正解はホモ接合)。

【0327】

図7の左は除外条件1を適用した後にデータについての信頼性値の分布図である。図7の右は除外条件2を適用した後にデータについての信頼性値の分布図である。図7に示すように除外条件1又は2の適用の結果、例外的事例数が有意に除外され妥当性の向上が認められた。

【0328】

＜試験例5＞異なるNGSターゲットパネルに対する妥当性の検証

本発明の妥当性を検証するため別途準備した16組データセットを用いて以下の検討を行った。試験例1に示した184SNPsターゲットパネルとは異なる、132SNPsのターゲットパネルの解析結果である。

母親の口腔粘膜サンプル、父親の口腔粘膜サンプル、母親plasmaサンプル及び新生児の口腔粘膜サンプルを分析したNGSによる遺伝子配列検査データを1組のデータセットとし、試験例1～3とは別途16組のデータセットを用意した。NGSは既知の132個のSNPsのある多型座位についてターゲットシーケンスを行ったものである。すなわち、用意したデータセットには、2,112個(16組×132個)のSNPsに関するデータが含まれている。

なお、本試験例で解析する132個のSNPsは、試験例1～3で解析した184個のSNPsに完全に重複するものではなく、71個のSNPsは試験例1～3で解析したSNPsとは別のSNPsである。

このデータセットのうち、父母ともにホモ接合として有するSNPsを抽出し531個のSNPsの信頼性値を算出した。

【0329】

16テストデータセットから算出した信頼性値(Fidelity)の分布図を図8に示す。

左は父母がそれぞれホモ接合で有している互いに異型のSNPs(胎児遺伝型の正解はヘテロ接合)と、父母がそれぞれホモ接合で有している互いに同型のSNPs(胎児遺伝型の正解はホモ接合)に関する信頼性値を集計したものである。

父母がそれぞれホモ接合で有している互いに異型のSNPsについては、176個のSNPs中175個で0.9以上の信頼性値を示した。また、父母がそれぞれホモ接合で有している互いに同型のSNPsについては、355個のSNPs中0.5以上の信頼性値を示すSNPが1個だけ存在した。

【0330】

右は試験例1とは異なるターゲットパネルのFidelity分布図を割合で表示したものである。

200テストデータセット(184SNPsのターゲットパネル)から算出されたモデル関数を、16テストデータセット(132ターゲットパネル)の解析に用いたが同様のFidelity分布図が得られた。

これらの結果より、本発明のFidelity算出方法は同じ試験システムを用いる場合、ターゲットパネルの種類に関係なく正確であることが確認された。

10

20

30

40

50

【0331】

<試験例6> 副次成分信号の真偽が不明なSNPsに対する妥当性の検証

試験例5で用いた16セットのデータの内、母がホモ接合で有している951SNPsのFidelity分布を新生児遺伝型がヘテロとホモに集計し図9にまとめた。

なお、図9で示した全てのSNPsは胎児Count Majorと胎児Count minorの合計300以上である。

【0332】

図9から明らかなように、父母の遺伝型を用いた胎児遺伝型の推定形は出生後に確定された子の遺伝型と一致した。新生児ホモSNPsの99.6% (575SNPsの内573SNPs)は0.2以下の低いFidelityを示し、新生児ヘテロSNPsの99.4% (376SNPsの内374SNPs)は0.8以上の高いFidelityを示した。

【0333】

従って、試験例1のモデル関数を用いたFidelity算出方法によれば、副次成分信号存在に対する真偽を示す父の遺伝型が分からない場合であっても、正確なFidelityの計算ができる。

これらの結果より、本発明の妥当性が確認された。

【0334】

<試験例7> モデル関数の作成 (その2)

試験例1で使用したものと同一のデータセットについて、母親及び父親ともにホモ接合である多型座位に関するものだけを抽出した。この抽出したデータセットに含まれる以下の表1に示す13因子について主成分分析を行った。表1に主成分分析の結果得られた第1主成分に関する固有ベクトルを示す。

【0335】

【表1】

	変 数	主成分1
(1)	胎児Count_major	0.008710919
(2)	胎児Count_minor	0.37133704
(3)	胎児Freq_major	-0.397335595
(4)	胎児Freq_minor	0.397377539
(5)	胎児Error	0.011977182
(6)	Count_Major / Ave.Error	0.012044526
(7)	Count_minor / Ave.Error	0.375885418
(8)	Frequency_Major / Ave.Error	0.001745322
(9)	Frequency_minor / Ave.Error	0.329854084
(10)	Forward read count_Major	0.007477871
(11)	Reverse read count_Major	0.009211129
(12)	MFAA count_Major	0.371226399
(13)	MFAA frequency_Major	0.397383319

【0336】

表1に示した13因子のうち(1)～(5)の内容は試験例1で説明した通りである。なお、表1における変数の表記として、「major」が含まれるものは主要成分信号に関するデータであり、「minor」が含まれるものは副次成分信号に関するデータである。また、表1における変数の表記として、「count」が含まれるものは信号強度に

関するデータであり、「freq」又は「frequency」が含まれるものは信号強度の割合に関するデータである。

すなわち、表1における変数の表記として「minor」及び「count」の両方が含まれる数値は、本発明でいう「副次成分信号強度」に該当する。

また、表1における変数の表記として「minor」及び「freq」若しくは「frequency」の両方が含まれる数値は、本発明でいう「副次成分混入率」に該当する。
【0337】

なお、表1の(7)は、前記特定の多型座位のアレルの存在を示す副次成分信号強度を、複数の多型座位におけるノイズの平均値により割った数値である。

また、表1の(9)は、前記特定の多型座位のアレルに起因する信号強度総和に対する前記副次成分信号強度の割合である副次成分混入率を、複数の多型座位におけるノイズの平均値により割った数値である。

【0338】

主成分分析により得られた第1主成分を基に試験例1と同様の手順によって、当該第1主成分 x_1 を説明変数、信頼性値を目的変数とするモデル関数 $f_1(x_1)$ を作成した。回帰分析の寄与率(R^2)は0.99以上と極めて良好であった。

この $f_1(x_1)$ と上述の $f_2(x_2)$ 及び $f_3(x_3)$ を乗じて、上記式4で表されるモデル関数 $f(x_1, x_2, x_3)$ を作成した。

【0339】

<試験例8>信頼性値の算出(その2)

試験例1で使用したものと同一のデータセットに含まれる表1に示す13因子について主成分分析を行った。主成分分析によって得られた第1主成分、副次成分信号強度絶対値及び副次成分混入率を試験例7で作成したモデル関数 $f(x_1, x_2, x_3)$ に入力し、信頼性値を算出した。5因子、又は13因子について主成分分析を行い算出した信頼性値(Fidelity)の分布図を図10に示す。

図10に示すように、本試験例でも、例外的結果がほぼ見られない極めて精度の高い結果が得られた。この結果より、試験例7で作成したモデル関数の妥当性と精度の高さが証明された。

【0340】

<試験例9>妥当性の検証(その2)

試験例6で用意したものと同一のデータセットを用意しデータセットに含まれる表1に示す13因子について主成分分析を行った。主成分分析によって得られた第1主成分、副次成分信号強度絶対値及び副次成分混入率を試験例7で作成したモデル関数 $f(x_1, x_2, x_3)$ に入力し、信頼性値を算出した。5因子、又は13因子について主成分分析を行い算出した信頼性値(Fidelity)の分布図を図11に示す。

図11に示すように、本試験例で副次成分信号存在に対する真偽を示す父の遺伝型が分からない場合も、例外的結果がほぼ見られない極めて精度の高い結果が得られた。この結果より、試験例7で作成したモデル関数の妥当性と精度の高さが証明された。

【産業上の利用可能性】

【0341】

本発明は出生前遺伝学的検査、癌スクリーニング検査、移植臓器定着モニタリング、感染症検査、法医学に応用できる。

10

20

30

40

50

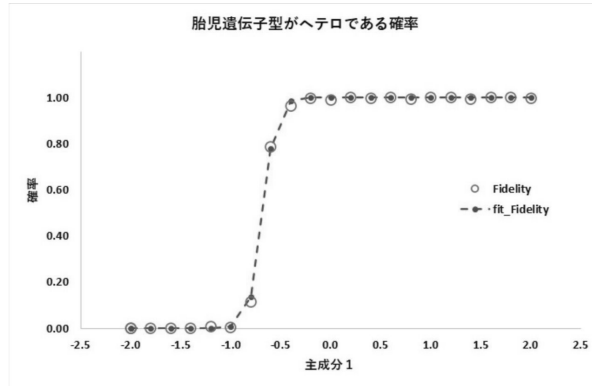
【要約】

本発明の解決すべき課題はc f f DNA、c t DNA、d d c f DNAのような副次核酸を微小な割合で含む混合核酸サンプルの分析データにおける、副次核酸の存在を示す信号の信頼性を評価する新規の技術を提供することである。

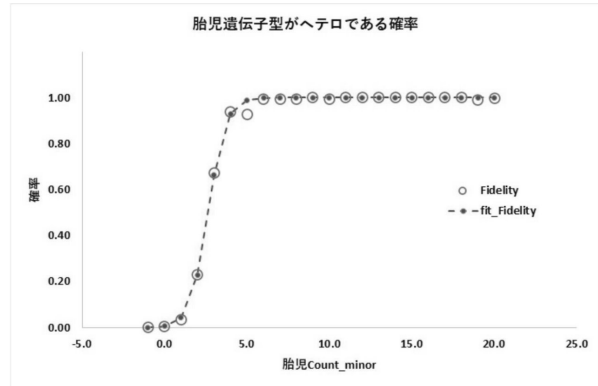
上記課題の解決手段は、分析データに含まれる少なくとも副次成分信号強度及び副次成分混入率を含む数値群について線形結合をした結果得られた合成変数と信頼性値について回帰分析を行い、信頼性値を算出するためのモデル関数を求めることである。

【図面】

【図 1】



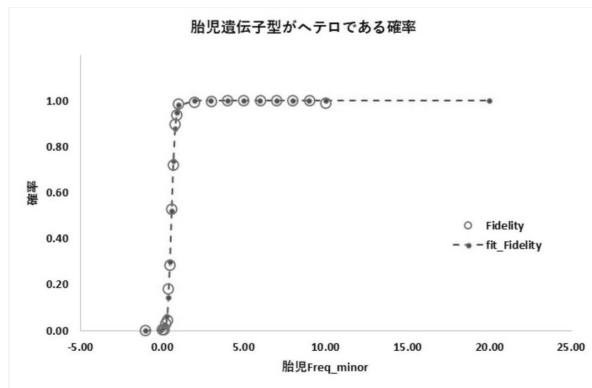
【図 2】



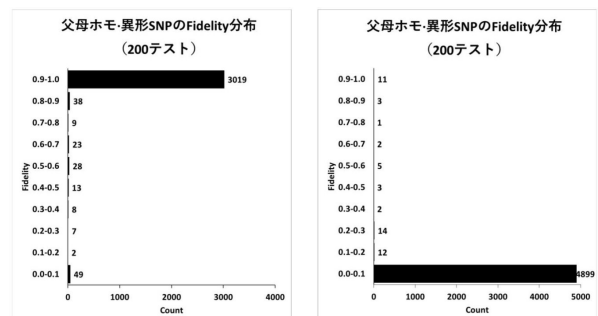
10

20

【図 3】



【図 4】

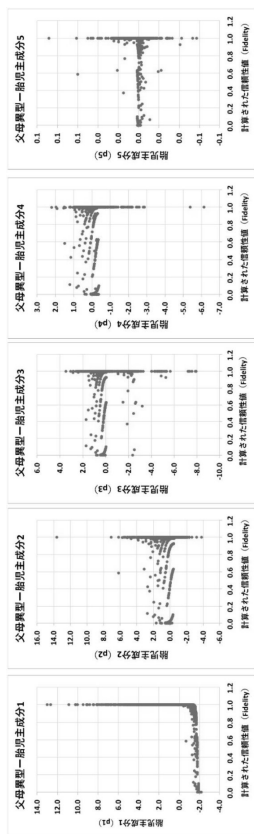


30

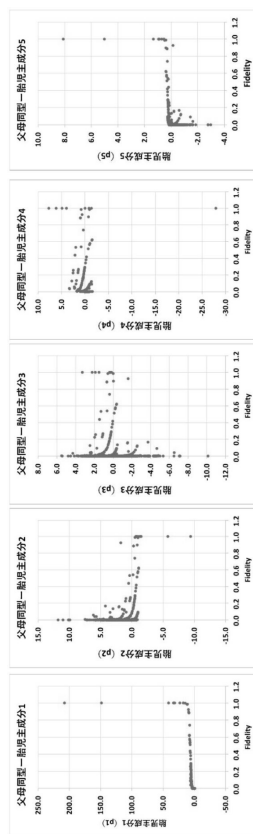
40

50

【図 5】



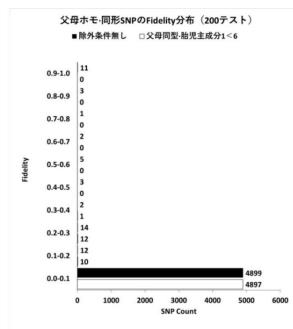
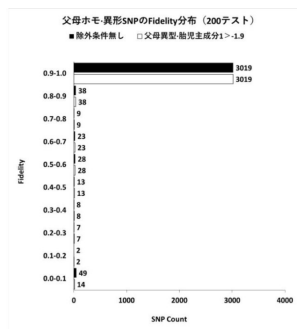
【図 6】



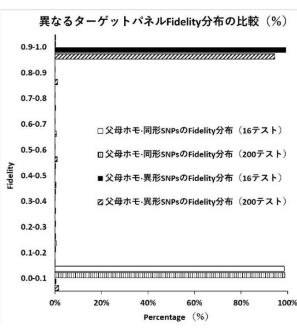
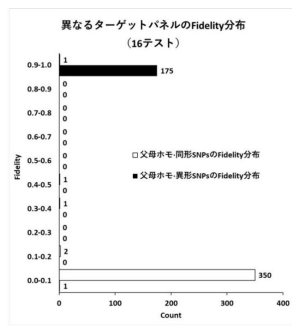
10

20

【図 7】



【図 8】

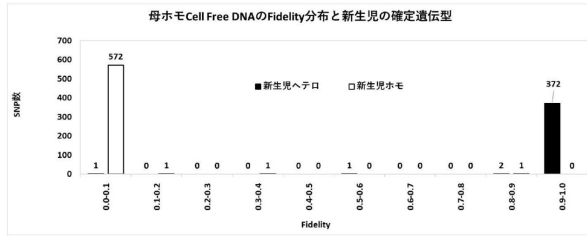


30

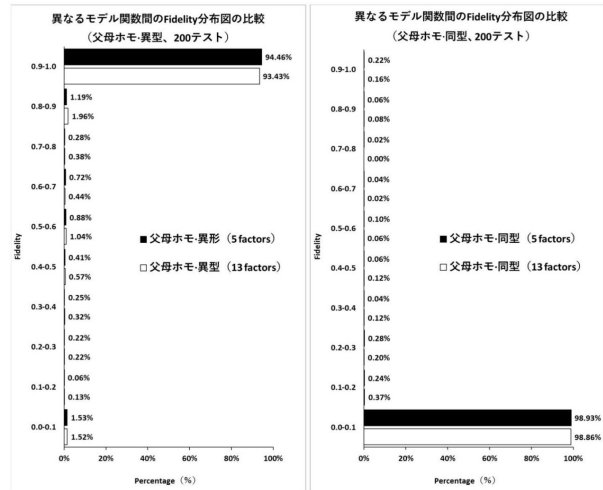
40

50

【図 9】

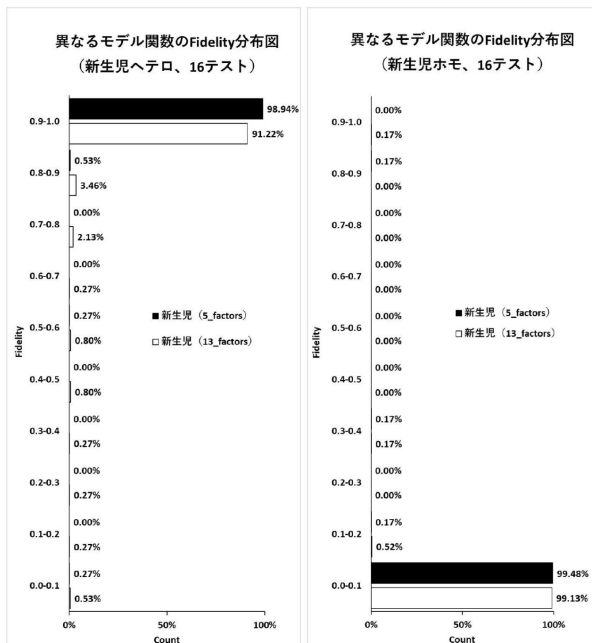


【図 10】



10

【図 11】



20

30

40

50

フロントページの続き

横浜研究所内

審査官 岡北 有平

(56)参考文献 特開 2016-034282 (JP, A)

特表 2014-502845 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

G16B 5/00 - 99/00