

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7124265号
(P7124265)

(45)発行日 令和4年8月24日(2022. 8. 24)

(24)登録日 令和4年8月16日(2022. 8. 16)

(51)Int. Cl. F I
G 1 6 B 40/00 (2019. 01) G 1 6 B 40/00

請求項の数 11 (全 22 頁)

(21)出願番号	特願2019-517725(P2019-517725)	(73)特許権者	503360115
(86)(22)出願日	平成30年5月11日(2018. 5. 11)		国立研究開発法人科学技術振興機構
(86)国際出願番号	PCT/JP2018/018393		埼玉県川口市本町四丁目1番8号
(87)国際公開番号	W02018/207925	(74)代理人	100137800
(87)国際公開日	平成30年11月15日(2018. 11. 15)		弁理士 吉田 正義
審査請求日	令和3年4月20日(2021. 4. 20)	(74)代理人	100148253
(31)優先権主張番号	特願2017-96060(P2017-96060)		弁理士 今枝 弘充
(32)優先日	平成29年5月12日(2017. 5. 12)	(74)代理人	100148079
(33)優先権主張国・地域又は機関	日本国(JP)		弁理士 梅村 裕明
		(74)代理人	100158241
			弁理士 吉田 安子
		(74)代理人	100168619
			弁理士 川村 昌子

最終頁に続く

(54)【発明の名称】 バイオマーカー検出方法、疾病判断方法、バイオマーカー検出装置、及びバイオマーカー検出プログラム

(57)【特許請求の範囲】

【請求項1】

生体に関する測定により得られたデータに基づいて、生体の状態を示すバイオマーカーを検出する、バイオマーカー検出方法であって、

1又は2以上の参照生体から得られたデータに基づく参照データセットを用意し、対象生体から得られた対象生体データを前記参照データセットに追加して対象データセットを生成し、

前記参照データセットの中の複数の因子項目の間で第1の相関係数を求め、

前記対象データセットの中の複数の因子項目の間で第2の相関係数を求め、

前記第1の相関係数と前記第2の相関係数との差分である差分相関係数を求め、

前記複数の因子項目のうちの一つの因子項目と、当該一つの因子項目に対して所定の相関係数を持つ1又は2以上の別の因子項目との間の前記差分相関係数の第1平均値を求め、前記参照データセットの中の前記複数の因子項目のうちの前記一つの因子項目に関する複数のデータの平均値を求め、前記対象生体データの中の前記一つの因子項目のデータと前記平均値との差を求め、(前記第1平均値×前記差)に基づく指標を求め、

前記指標に基づいてバイオマーカーを選択する

工程を含む、バイオマーカー検出方法。

【請求項2】

前記指標を求める工程が、前記別の因子項目と、当該別の因子項目に対して所定の相関係数を持つ1又は2以上の更に別の因子項目との間の前記差分相関係数の第2平均値を求

める工程を更に含み、

前記指標が、 $(\text{前記差} \times \text{前記第 1 平均値}) / (\text{前記第 2 平均値})$ で表される式により求められる、請求項 1 に記載のバイオマーカー検出方法。

【請求項 3】

前記指標を求める工程が、前記複数のデータの**前記平均値**を求める工程と、前記差を求める工程と、前記第 1 平均値を求める工程と、を前記複数の因子項目のすべてに対して行う工程を更に含み、

前記複数の因子項目のすべてに対して、前記指標が、前記第 1 平均値及び前記差に基づいて求められる、請求項 1 に記載のバイオマーカー検出方法。

【請求項 4】

前記指標を求める工程が、

前記複数のデータの**前記平均値**を求める工程と、

前記差を求める工程と、前記第 1 平均値を求める工程と、を前記複数の因子項目のすべてに対して行う工程と、

前記別の因子項目と、当該別の因子項目に対して所定の相関係数を持つ前記 1 又は 2 以上の更に別の因子項目との間の**前記差分相関係数の前記第 2 平均値**を求める工程と、

を前記複数の因子項目のすべてに対して行う工程を更に含み、

前記複数の因子項目のすべてに対して、前記指標が、 $(\text{前記差} \times \text{前記第 1 平均値}) / (\text{前記第 2 平均値})$ に基づいて求められる、

請求項 2 に記載のバイオマーカー検出方法。

【請求項 5】

前記バイオマーカーを選択する工程が、

前記複数の因子項目のすべてに対して求められた前記指標を大きい順に並べたときに、大きい順に 1 又は 2 以上の所定の数の指標を前記バイオマーカーとして選択する、請求項 3 又は 4 に記載のバイオマーカー検出方法。

【請求項 6】

前記 1 又は 2 以上の参照生体から得られたデータは、良好な健康状態にある生体から取得したデータである、請求項 1 から 5 のいずれか一項に記載のバイオマーカー検出方法。

【請求項 7】

前記 1 又は 2 以上の参照生体から得られたデータは、特定の病気に罹っている生体から取得したデータである、請求項 1 から 5 のいずれか一項に記載のバイオマーカー検出方法。

【請求項 8】

前記対象データセットを生成する工程が、前記対象生体から複数回得られた対象生体データの中の**前記複数の因子項目のデータを前記参照データセットに追加して前記対象データセットを生成する**、請求項 1 から 7 のいずれか一項に記載のバイオマーカー検出方法。

【請求項 9】

請求項 1 から 8 のいずれか一項に記載のバイオマーカー検出方法を用いて前記バイオマーカーを検出し、検出された前記バイオマーカーと、あらかじめ特定された所定の疾病に対応したバイオマーカーの遺伝子群とを比較することにより、検出された前記バイオマーカーが前記所定の疾病に対応したバイオマーカーであるか否かを判断する、疾病判断方法。

【請求項 10】

生体に関する測定により得られた複数の因子項目のデータに基づいて、検査対象である対象生体の状態の指標となるバイオマーカーを検出する、バイオマーカー検出装置であって、

1 又は 2 以上の参照生体から得られた参照生体データの中の複数の因子項目のデータに基づく参照データセットを記憶する第 1 の記憶部と、

前記対象生体から得られた対象生体データの中の前記複数の因子項目のデータを前記参照データセットに追加して対象データセットを生成し、記憶する第 2 の記憶部と、

10

20

30

40

50

前記第 1 の記憶部に記憶される前記参照データセットの中の前記複数の因子項目の間で第 1 の相関係数を求め、前記第 2 の記憶部に記憶される前記対象データセットの中の前記複数の因子項目の間で第 2 の相関係数を求め、前記第 1 の相関係数と前記第 2 の相関係数との差分である差分相関係数を求め、前記複数の因子項目のうちの一つの因子項目と、当該一つの因子項目に対して所定の相関係数を持つ 1 又は 2 以上の別の因子項目との間の前記差分相関係数の第 1 平均値を求め、前記参照データセットの中の前記複数の因子項目のうちの前記一つの因子項目に関する複数のデータの平均値を求め、前記対象生体データの中の前記一つの因子項目のデータと前記平均値との差を求め、(前記第 1 平均値 × 前記差)に基づく指標を求め、前記指標に基づいてバイオマーカーを選択する制御部と、
を備える、バイオマーカー検出装置。

10

【請求項 11】

請求項 1 から 10 のいずれか一項に記載のバイオマーカー検出方法を、生体に関する測定により得られた複数の因子項目のデータに基づいて、検査対象である対象生体の状態の指標となるバイオマーカーを検出する、バイオマーカー検出装置であって、前記 1 又は 2 以上の参照生体から得られた参照生体データの中の複数の因子項目のデータに基づく参照データセットを記憶する第 1 の記憶部と、前記対象生体から得られた対象生体データの中の前記複数の因子項目のデータを前記参照データセットに追加して対象データセットを生成し、記憶する第 2 の記憶部と、前記第 1 の記憶部に記憶される前記参照データセットの中の前記複数の因子項目の間で第 1 の相関係数を求め、前記第 2 の記憶部に記憶される前記対象データセットの中の前記複数の因子項目の間で第 2 の相関係数を求め、前記第 1 の相関係数と前記第 2 の相関係数との差分である差分相関係数を求め、前記複数の因子項目のうちの一つの因子項目と、当該一つの因子項目に対して所定の相関係数を持つ 1 又は 2 以上の別の因子項目との間の前記差分相関係数の第 1 平均値を求め、前記参照データセットの中の前記複数の因子項目のうちの前記一つの因子項目に関する複数のデータの平均値を求め、前記対象生体データの中の前記一つの因子項目のデータと前記平均値との差を求め、(前記第 1 平均値 × 前記差)に基づく指標を求め、前記指標に基づいてバイオマーカーを選択する制御部と、を備えるバイオマーカー検出装置に実行させる命令を含むバイオマーカー検出プログラム。

20

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、生体に対する測定により得られた因子項目のデータに基づいて、生体の状態を示すバイオマーカーを検出するバイオマーカー検出方法、疾病判断方法、バイオマーカー検出装置、及びバイオマーカー検出プログラムに関する。

【背景技術】

【0002】

生体の状態(例えば健康状態等)は、気候システム、生態システム、経済システム等の複雑システムと同様に、ある過渡点(分岐点)を超えると、良好又は正常(以下、単に「良好」と呼ぶ)な状態から疾病状態に急激に変化することが知られている(特許文献 1 及び非特許文献 1 ~ 5)。また、複雑システムでは、過渡点において、揺らぎが増大するとともに、その揺らぎが強く相関する変数が存在することが知られている。その一因として、システムを安定に維持するための回復力が低下し、外乱に対し脆弱となっていることが挙げられる。

40

【0003】

以下、図 1 を参照しながら、人が病気になる場合を例に説明する。図 1 は、疾病の進行プロセスを示す概略図である。図 1 において、縦軸は健康状態を概念的に表し、矢印と反対の方向(下向き)に向かうほど健康状態が悪くなることを示している。また、横軸は時間を表している。図中の点 a は、ある人の健康状態が良好であることを示している。この場合、本人には自覚症状がなく、従来の健康診査での検査結果にも特段の異常は見られない。図 1 に示すように、時間の経過とともに健康状態が悪化しても、その変化は極

50

めて緩やかなため、依然として、体調の変化を感じることはない。しかし、緩やかな悪化が続き、図中の点 b で示す過渡的な状態（健康と病気の間の状態）に達すると、健康状態は急激（あるは緩慢な場合もある）に悪化し、早期疾病状態に至る（点 c）。この状態に至ると、本人も体の不調を感じることができるようになり、やがて疾病状態（点 d）に陥る。ここにいう過渡的な状態（点 b）は、本人には特段の自覚症状がないが、直ちにも早期疾病状態に移行してしまう可能性が高い状態である。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】国際公開第 2014/050160号

10

【非特許文献】

【0005】

【非特許文献 1】Jose G. Venegas, Tilo Winkler, Guido Musch, Marcos F. Vidal Melo, Dominick Layfield, Nora Tgavalekos, Alan J. Fischman, Ronald J. Callahan, Giacomo Bellani, and R. Scott Harris, "Self organized patchiness in asthma as a prelude to catastrophic shifts," Nature 434, Nature Publishing Group, pp. 777-782 (2005)

【非特許文献 2】Patrick E. McSharry, Leonard A. Smith, and Lionel Tarassenko, "Prediction of epileptic seizures: are nonlinear methods relevant?," Nature Medicine 9, Nature Publishing Group, pp. 241-242 (2003)

20

【非特許文献 3】Roberto Pastor Barriuso, Eliseo Guallar, and Josef Coresh, "Transition models for change point estimation in logistic regression," Statistics in Medicine 22(7), Wiley Blackwell, pp. 1141-1162 (2003)

【非特許文献 4】Paek SH et al. "Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma," Cancer 104, Wiley Blackwell, pp. 580-590 (2005)

【非特許文献 5】Liu, J.K., Rovit, R.L., and Couldwell, W.T., "Pituitary Apoplexy," Seminars in Neurosurgery 12, Thieme, pp. 315-320 (2001)

【発明の概要】

【発明が解決しようとする課題】

30

【0006】

ところで、人が病院へ行くのは、多くの場合、体調不良を自覚してから、すなわち、過渡的な状態（点 b）を過ぎて早期疾病状態（点 c）や疾病状態（点 d）に至ってからである。病院では、検査や画像診断などが行なわれる。このような診断では、殆どの場合、検査や画像診断で異常が見つかったときには、診察を受けた人が既に病気に罹っている。

一方、上述の過渡的な状態（点 b）を把握（早期発見）することができれば、この時点で治療することにより（早期治療）、早期疾病状態（点 c）に達するのを防ぐことができ（つまり、病気にならなくて済み）、良好な健康状態（点 a）を維持できることが推測される。

40

【0007】

本発明は、上記の事情に鑑み、疾病状態へ遷移する前の過渡的な状態（健康と病気の間の状態）を検出することができる、バイオマーカー検出方法、疾病判断方法、バイオマーカー検出装置、及びバイオマーカー検出プログラムを提供する。

【課題を解決するための手段】

【0008】

本発明の第 1 の態様は、生体に関する測定により得られたデータに基づいて、生体の状態の指標となるバイオマーカーを検出するバイオマーカー検出方法を提供する。このバイオマーカー検出方法は、1 又は 2 以上の参照生体から得られたデータに基づく参照データセットを用意し、対象生体から得られた対象生体データを参照データセットに追加して対

50

象データセットを生成し、参照データセットの中の複数の因子項目の間で第1の相関係数を求め、対象データセットの中の複数の因子項目の間で第2の相関係数を求め、第1の相関係数と第2の相関係数との差分である差分相関係数を求め、複数の因子項目のそれぞれについて差分相関係数に基づく指標を求め、指標に基づいてバイオマーカーを選択する工程を含んでいる。

【0009】

本発明の第2の態様は、生体に関する測定により得られた複数の因子項目のデータに基づいて、検査対象である対象生体の状態の指標となるバイオマーカーを検出する、バイオマーカー検出装置を提供する。このバイオマーカー検出装置は、1又は2以上の参照生体から得られた参照生体データの中の複数の因子項目のデータに基づく参照データセットを記憶する第1の記憶部と、対象生体から得られた対象生体データの中の複数の因子項目のデータを参照データセットに追加して対象データセットを生成し、記憶する第2の記憶部と、第1の記憶部に記憶される参照データセットの中の複数の因子項目の間で第1の相関係数を求め、第2の記憶部に記憶される対象データセットの中の複数の因子項目の間で第2の相関係数を求め、第1の相関係数と第2の相関係数との差分である差分相関係数を求め、複数の因子項目のそれぞれについて差分相関係数に基づく指標を求め、指標に基づいてバイオマーカーを選択する制御部とを含んでいる。

10

【0010】

本発明の第3の態様は、上述のバイオマーカー検出方法を上述のバイオマーカー検出装置に実行させる命令を含むバイオマーカー検出プログラムを提供する。

20

【発明の効果】

【0011】

本発明の実施形態によれば、次の効果を得ることが出来る。(1)疾病状態へ遷移する前の過渡的な状態を示す指標(新しいバイオマーカー)を特定できる、また、(2)その指標を使って早期診断ができる。

【図面の簡単な説明】

【0012】

【図1】良好な健康状態から疾病状態への移行プロセスを説明する模式図である。

【図2】本発明の第1の実施形態によるバイオマーカーの検出方法を説明するフローチャートである。

30

【図3】第1の実施形態のバイオマーカーの検出方法を説明する模式図である。

【図4】図3に引き続き、第1の実施形態のバイオマーカーの検出方法を説明する模式図である。

【図5】第1の実施形態のバイオマーカーの検出方法を検証するために行った実験の結果を示すグラフである。

【図6】本発明の第2の実施形態によるバイオマーカーの検出装置の構成を示すブロック図である。

【図7】本発明の第3の実施形態によるバイオマーカーの検出方法を説明するフローチャートである。

【図8】本発明の第3の実施形態によるバイオマーカーの検出方法を説明する模式図である。

40

【図9】本発明の第4の実施形態によるバイオマーカーの検出方法を説明する模式図である。

【発明を実施するための形態】

【0013】

以下、添付の図面を参照しながら、本発明の限定的でない例示の実施形態について説明する。添付の全図面中、同一または対応する部材または部品については、同一または対応する参照符号を付し、重複する説明を省略する。

(第1の実施形態)

以下、図2から図5までを参照しながら、本発明の第1の実施形態によるバイオマーカー

50

一の検出方法を説明する。図2は、第1の実施形態によるバイオマーカーの検出方法を説明するフローチャートであり、図3及び図4は、第1の実施形態によるバイオマーカーの検出方法を示す模式図である。

第1の実施形態によるバイオマーカーの検出方法は、図2に示すように、参照データセットを用意するステップ(ST1)、検査対象データセットを生成するステップ(ST2)、第1の相関係数を計算するステップ(ST3)、第2の相関係数を計算するステップ(ST4)、差分相関係数を計算するステップ(ST5)、指標を求めるステップ(ST6)、バイオマーカーを選択するステップ(ST7)を含んでいる。

【0014】

<参照データセットを用意するステップ(ST1)>

初めに参照データセットDr(図3)が用意される。参照データセットDrは、バイオマーカーの検出のための参照用のデータとして用いられるデータの集合である。参照データセットDrは、複数の生体データ提供者S1、S2、・・・、Snの各々に対し生体に関する測定、すなわち生体に対して測定を行うことにより用意される。生体データ提供者S1、S2、・・・、Snには、通常は第1の実施形態によるバイオマーカーの検出方法の検査対象者Se(以下、被検者Seと呼ぶ)は含まれないが、含まれる場合も考えられる。

また、第1の実施形態においては、生体データ提供者S1、S2、・・・、Snは良好な健康状態にあり、これらの人々から生体サンプル(例えば血液)が採取される。ここで、良好な健康状態にある生体データ提供者とは自覚症状がない人であって良く、健康診断などにより健康と判断されることは要しない。ただし、特定の病気に罹っている可能性が高いことが判明している人は、生体データ提供者にはなれない。また、生体データ提供者S1、S2、・・・、Snの人数は、特に限定されることはないが、例えば少なくとも8人であると好ましく、例えば10人以上であると更に好ましい。

【0015】

各生体サンプルに対して測定かつ/又は所定の処理が行われ、第1の実施形態においては、因子項目としての遺伝子g1、g2、・・・、gmに関するデータD(発現量)が取得される。この場合、生体サンプルは、DNAチップなどのハイスループット技術により処理されることが好ましい。ハイスループット技術によれば、1つの生体サンプルから2万個以上の遺伝子の発現量を測定することが可能である。取得された遺伝子発現量のデータは、説明の便宜上、図3に示すようにマトリックス状にまとめられるものとする。このマトリックス状の参照データセットDr内の各四角()が、生体データ提供者S1の遺伝子g1の発現量データ、遺伝子g2の発現量データ、・・・、遺伝子gmの発現量データを模式的に示している。

【0016】

<検査対象データセットを生成するステップ(ST2)>

次に、被検者Seからも同様に生体サンプルが採取され、遺伝子g1、g2、g3、・・・、gmに関する被検者データセットDeが取得される(図3)。その後、先に用意された参照データセットDrが複製され、複製された参照データセットDrに被検者データセットDeを追加することにより、検査対象データセットDcが作成される。

【0017】

<第1の相関係数を計算するステップ(ST3)>

次いで、参照データセットDr内のデータ間の相関係数が計算される。図3から分かるように、参照データセットDrにおいては、遺伝子g1について、生体データ提供者S1、S2、・・・、Snのn人に対応したn個のデータが存在している。同様に、遺伝子g2についてもn個のデータが存在している。まず、遺伝子g1の集合CL1rと、遺伝子g2の集合CL2rとの間で相関係数が計算される。第1の実施形態では、2つの集合の間でピアソンの積率相関係数(PPC)が計算される。続けて、遺伝子g1の集合と、遺伝子g3、g4、・・・、gmの各集合との間でも、それぞれPPCが得られる。

さらに続けて、遺伝子g2の集合と、遺伝子g3、g4、・・・、gmの各集合との間

10

20

30

40

50

でも P C C が得られ、遺伝子 g 3 の集合と、遺伝子 g 4、g 5、・・・、g m の各集合との間でも P C C が得られる。このようにして、遺伝子 g 1、g 2、g 3、・・・、g m の集合のうちのある一対の集合の間で P C C が得られる。なお、説明の便宜上、参照データセット D r について得られた相関係数 P C C を P C C r と呼ぶ。

【 0 0 1 8 】

< 第 2 の相関係数を計算するステップ (S T 4) >

次に、検査対象データセット D c 内のデータ間の相関係数が計算される。図 3 に示すように、検査対象データセット D c では各遺伝子の集合に属するデータの数は n + 1 となるが、参照データセット D r についての計算方法と同じ方法が用いられ、検査対象データセット D c について P C C が計算される。説明の便宜上、検査対象データセット D c における相関係数 P C C を P C C c と呼ぶ。

10

【 0 0 1 9 】

< 差分相関係数を計算するステップ (S T 5) >

この後、上述の相関係数 P C C r 及び P C C c の差分相関係数が計算される。すなわち、参照データセット D r における遺伝子 g 1 の集合と遺伝子 g 2 の集合との間の相関係数 P C C r と、検査対象データセット D c における遺伝子 g 1 の集合と遺伝子 g 2 の集合との間の相関係数 P C C c との差分相関係数 P C C ((P C C r - P C C c) の絶対値) が計算される。次いで、遺伝子 g 1 (の集合) と遺伝子 g 3 (の集合) との相関係数に関して、参照データセット D r と検査対象データセット D c との間で差分相関係数 P C C が計算される。続けて、遺伝子 g 1 と、遺伝子 g 4、g 5、・・・、g m のそれぞれとの相関係数に関し、参照データセット D r と検査対象データセット D c との間で差分相関係数 P C C (図 3) が計算される。同様に、遺伝子 g 2 と、遺伝子 g 3、g 4、・・・、g m のそれぞれとの相関係数に関しても、データセット D r、D c の間で差分相関係数 P C C が計算される。このようにして、参照データセット D r と検査対象データセット D c の間で、2 組の遺伝子の集合ごとの対応する組み合わせについて、差分相関係数 P C C が計算される。

20

【 0 0 2 0 】

次に、このようにして計算された P C C のうち、閾値 (固定値であっても変数値であっても良い) を超えた P C C が抽出される。ここで、説明の便宜上、選択された P C C に基づいて図 4 に示す差分相関係数ネットワーク N r c が構築されるものとする。例えば、遺伝子 g 1 の相関係数 P C C と遺伝子 g 2 の相関係数 P C C との間の差分相関係数 P C C が抽出されたとすると、遺伝子 g 1 に対応するノード G 1 と、遺伝子 g 2 に対応するノード G 2 とが設定され、これらのノード G 1、G 2 が枝 B で接続されている (リンクされる)。同様の操作が、抽出されたすべての差分相関係数 P C C について行われ、差分相関係数ネットワーク N r c が構築される (図 4)。

30

【 0 0 2 1 】

< 指標を求めるステップ (S T 6) >

次に、ノード (結節点) のそれぞれについて、他のノードとの相関係数の平均値が計算される。まず、着目するノードと、そのノードについての 1 次のノードとの間の相関係数の平均値 s P C C i n が計算される。ここで 1 次のノードとは、着目するノードにリンクするノードである。例えば、図 4 のノード G 1 に着目すると、ノード G 2、G 3、及び G 4 がノード G 1 にリンクしている。よって、これらのノード G 2、G 3、及び G 4 がノード G 1 の 1 次のノードである。ノード G 1 及び G 2 の間の相関係数と、ノード G 1 及び G 3 の間の相関係数と、ノード G 1 及び G 4 の間の相関係数とが平均化され、平均値 s P C C i n が得られる。

40

【 0 0 2 2 】

なお、例えばノード G 1 及び G 2 の間の相関係数は、先に計算された遺伝子 g 1 及び g 2 の間の差分相関係数 (P C C) に相当する。すなわち、この例における平均値 s P C C i n は、遺伝子 g 1 及び g 2 の間の P C C と、遺伝子 g 1 及び g 3 の間の P C C と、遺伝子 g 1 及び g 4 の間の P C C との平均値である。

50

また、図4のノードG5に着目した場合、ノードG5は1つのノードG2のみがリンクしている。この場合は平均ではなく、ノードG5及びG2の間の相関係数を $sPCCin$ として用いても良い。

【0023】

次いで、着目するノードに関する1次のノードと2次のノードとの間の相関係数の平均値 $sPCCout$ が計算される。ここで2次のノードとは、着目するノードの1次のノードにリンクするノードであって、着目するノードにリンクしていないノードである。図4のノードG1に関しては、1次のノードG2にリンクするノードG5、G6、及びG7と、1次のノードG3にリンクするノードG8と、1次のノードG4に直接にリンクするノードG9及びG10とが2次のノードに相当する。よって、1次のノードG2と2次のノードG5、G6、及びG7のそれぞれとの間、1次のノードG3と2次のノードG8の間、1次のノードG4と2次のノードG9及びG10のそれぞれとの間の合計6つの相関係数が平均化され、平均値 $sPCCout$ が得られる。

10

ここでも、例えばノードG2及びG5の間の相関係数は、これらのノードにそれぞれ対応する遺伝子 $g2$ 及び $g5$ の間の差分相関係数 PCC に相当する。

【0024】

ここで再び図3を参照する。グラフGRにおいて、横軸が遺伝子の発現量を示し、縦軸が確率を示している。グラフGRには、参照データセットDr中の(すなわち、複数の生体データ提供者 $S1$ 、 $S2$ 、 \dots 、 Sn の) 遺伝子 $g1$ についての発現量 - 確率曲線 Cr が破線で示されている。複数の生体データ提供者 $S1$ 、 $S2$ 、 \dots 、 Sn の遺伝子 $g1$ の発現量の確率はグラフGRに示すように、ほぼ正規に分布し、この分布図から発現量の平均値 μ_r が求められる。また、グラフGRには被検者 Se の遺伝子 $g1$ のデータ(発現量)も実線で示されており、その発現量を μ_e とすると、 $|\mu_r - \mu_e|$ により平均値 μ_r との差 sED が得られる。

20

【0025】

次に、以下の式(1)に従って、上記の平均値 $sPCCin$ と、平均値 $sPCCout$ と、差 sED とから、ノードG1(すなわち、遺伝子 $g1$) についての指標(スコア) Is が求められる。

$$Is = sED \times sPCCin / sPCCout \quad \dots \quad \text{式(1)}$$

次いで、指標 Is は、すべてのノードG2、G3、 \dots 、G10について同様に計算される。

30

【0026】

<バイオマーカーを選択するステップ(ST7)>

次に、図4に模式的に示すように、すべてのノードについて計算された複数の指標 Is が大きい順にソートされる。ソートされた指標 Is から大きい順に所定の数の指標 Is が選択される。これにより、選択された所定の数の指標 Is に対応する遺伝子がバイオマーカーとして検出されたこととなる。なお、図4に示すように閾値 Vth を定め、閾値 Vth を超える指標 Is が選択されるようにしても良い。また、すべてのノードについて計算するのではなく、あらかじめ所定の sED 値(例えばある値以上の sED) を持つノードを選択し、その選択されたノードについて PCC を計算し、指標を計算しても良い。

40

【0027】

以上のように、第1の実施形態のバイオマーカーの検出方法によれば、まず、生体データ提供者 $S1$ 、 $S2$ 、 \dots 、 Sn から得られた参照データセットDrの因子項目の間の相関係数 $PCCr$ が計算され、参照データセットDrに被検者 Se の被検者データセットDeを加えることにより得られた検査対象データセットDcの各因子項目の間の相関係数 $PCCc$ が計算される。次に、これらの差分相関係数 PCC が求められ、その中から閾値を超える差分相関係数 PCC が抽出され、これにより差分相関係数ネットワーク Nrc が構築される。差分相関係数ネットワーク Nrc のすべてのノードのそれぞれについて、 $sPCCin$ と $sPCCout$ が計算される。一方、各因子項目についての参照データセットDrの平均値と検査対象データセットDcとの差 sED が求められる。 sED と s

50

PCCinとsPCCoutを含む式(1)により指標Isが計算され、閾値Vthを超える指標Isに対応する遺伝子がバイオマーカーとして検出される。

【0028】

バイオマーカーが検出されるのは、被検者Seが、(見かけからは分からないものの)健康状態の点で、生体データ提供者S1、S2、・・・、Sn(の平均)と相違しているからである。このことは、仮に、被検者Seが、生体データ提供者同様に良好な健康状態(点a)にあれば、データセットDr、Dc間の差分相関係数PCCはほぼゼロになり、よってバイオマーカーは検出されないことから明らかである。

一方、先に説明した過渡的な状態(図1の点b)は、被検者自身には特段の自覚症状がないが、直ちにでも早期疾病状態に移行してしまう可能性のある状態である。このような状態においては、一般に、揺らぎが増大するとともに、その揺らぎが強く相関する変数が存在することが知られている。参照データセットDrと検査対象データセットDcとの間で大きな差異を示す因子項目は、過渡的な状態において、値の振れ幅が大きくなり、その動きが強く相関する生体的な因子に相当すると考えられる。したがって、バイオマーカーが検出されたことをもって、被検者Seの健康状態が過渡的な状態にあると判断することができる。

【0029】

次に、第1の実施形態によるバイオマーカーの検出方法の効果を検証するために行った実験について説明する。この実験では、健康な被検者Se1~Se17の17名に対し、インフルエンザウィルスを鼻腔内接種した後、第1の実施形態によるバイオマーカーの検出方法を時間の経過とともに15回行った。図5は、その実験の結果を示す表である。表中の「N」は、第1の実施形態によるバイオマーカーの検出方法によりバイオマーカーが検出されなかったことを示し、「C」は、第1の実施形態によるバイオマーカーの検出方法によりバイオマーカーが検出されたことを示し、「S」は、被検者が実際にインフルエンザに罹ったことを示している。

この表から、バイオマーカーが検出されなかった被検者Se2、Se3などは、疾病状態にはならず、バイオマーカーが検出されたすべての被検者Se1、Se5などが、その後、疾病状態になっていることが分かる。この結果から、第1の実施形態によるバイオマーカーの検出方法の有益な効果は明らかである。

【0030】

なお、この実験では、インフルエンザウィルスを被検者に鼻腔内接種した後に、バイオマーカーの検出を行ない、バイオマーカーが検出された被検者がインフルエンザを発病したことから、インフルエンザ発症の予兆がバイオマーカーの検出により見出されたといえることができる。しかし、一般には、バイオマーカーが検出されたことから分かるのは、健康状態が過渡的であること、または何らかの病気を発症する予兆があることであり、具体的な病気を特定することはできない。前述したインフルエンザの実験のように、対応する因子項目と病気との対応付けがなされれば、バイオマーカーが検出された時点で、後に発症する可能性がある病気まで特定することが可能となる。すなわち、第1の実施形態にかかるバイオマーカーは、特定の病気を発症する予兆となり得るといって利点を有している。

【0031】

(第2の実施形態)

次に、本発明の第2の実施形態によるバイオマーカーの検出装置について説明する。図6は、第2の実施形態による検出装置の構成例を示すブロック図である。図6に示すバイオマーカーの検出装置1は、制御部10、記憶部11、記憶部12、入力部13、出力部14、及びインタフェイス部15を備えている。

【0032】

制御部10は、CPU(Central Processing Unit)などの回路を用いて構成され、検出装置1全体を制御する。記憶部11は、HDD(Hard Disk Drive)などの磁気記録装置、SSD(Solid State Disk)などの不揮発性の補助記録装置を含む。記憶部11には、本発明の実施形態による検出プログラム11aを始めとする各種プログラムが記録され

10

20

30

40

50

ている。検出プログラム 11a は、第 1 の実施形態によるバイオマーカーの検出方法を検出装置 1 に実施させる命令（コード）を含んでいる。

【0033】

なお、検出プログラム 11a は、コンピュータ可読記憶媒体 13a に格納することができ、検出プログラム 11a を例えば入力部 13 を通してコンピュータ可読記憶媒体 13a から記憶部 11 にインストールすることができる。ここで、コンピュータ可読記憶媒体 13a は、ハードディスクドライブ（HDD）、固体ドライブ（SSD）、ランダムアクセスメモリ（RAM）、読出専用メモリ（ROM）、消去可能プログラム可能 ROM（EPROM 又はフラッシュメモリ素子）、コンパクトディスク ROM（CD-ROM）やデジタル多用途ディスク ROM（DVD-ROM）などの光ディスク記憶メディア、磁気記憶メディアなどを含む非一時的又は有形のコンピュータ可読記憶媒体であっても良い。

10

【0034】

また、記憶部 11 は、第 1 の実施形態にて説明した参照データセット Dr を記録することができる。参照データセット Dr は、入力部 13 を通して記憶部 11 に入力され、記録されても良いし、インタフェイス部 15 を通して、例えば外部のデータベース等から入力され、記録されても良い。

【0035】

記憶部 12 は、SDRAM（Synchronous Dynamic Random Access Memory）、SRAM（Static Random Access Memory）などの揮発性の記憶装置を含む。記憶部 12 は、第 1 の実施形態にて説明した被検者データセット De（図 3）を記憶することができる。被検者データセット De は、好ましくは入力部 13 を通して記憶部 12 に記憶されても良い。なお、第 1 の実施形態にて説明した検査対象データセット Dc（図 3）は、例えば、記憶部 11 に記録されている参照データセット Dr が制御部 10 へ読み出され、記憶部 12 に記憶されている被検者データセット De が制御部 10 へ読み出され、制御部 10 にて被検者データセット De を参照データセット Dr に追加することにより生成され得る。生成された検査対象データセット Dc を記憶部 12 に記憶しても良い。

20

【0036】

入力部 13 は、キーボードやマウス、入出力（I/O）装置などのハードウェア、及びドライバなどのソフトウェアを備える。I/O 装置はコンピュータ可読記憶媒体 13a にアクセスすることができる。出力部 14 は、ディスプレイやプリンタなどのハードウェア、及びドライバなどのソフトウェアを備える。

30

【0037】

インタフェイス部 15 は、外部から各種データを取得する。具体的には、インタフェイス部 15 は、通信網を介してデータを取り込むため、LAN（Local Area Network）ケーブルが接続されるポートや、測定機器に接続可能なパラレルケーブルなどの専用線が接続されるポートなどのハードウェア、及びドライバなどのソフトウェアを備える。

【0038】

記憶部 11 に記録されている検出プログラム 11a を、記憶部 12 に記憶し、制御部 10 の制御に基づいて実行することにより、第 2 の実施形態による検出装置 1 として機能する。

40

なお、第 2 の実施形態では記憶部 11 と記憶部 12 は個別に構成されているが、他の実施形態においては、一つのハードウェアで構成されていても良い。すなわち、一つのハードウェアのなかの異なる領域を記憶部 11 及び記憶部 12 として用いても構わない。

【0039】

なお、検出装置 1 を構成する制御部 10、記憶部 11、記憶部 12、入力部 13、出力部 14、及びインタフェイス部 15 は、単一の筐体内又は同一箇所に設置されてもよいし、これらの構成要素の各々又は一部を別の箇所の一又は複数の筐体内に設置し、有線又は無線のネットワークを介して接続するようにしてもよい。また、検出装置 1 によって実行される各種処理は、インターネット等のネットワークを介してクラウドコンピューティングを用いて実行されるようにしてもよい。

50

【 0 0 4 0 】

また、複数台の検出装置 1 を設置し、上述のバイオマーカの検出方法を含む各種処理を並行して実行するようにしてもよい。あるいは、検出装置 1 の構成要素の各々を複数のユニットにより構成するようにし（例えば、2 つの制御部 1 0、2 つの記憶部 1 1 など）、各種処理を並行して実行するようにしてもよい。このような並列処理により、高速化を図ることができ、また、データを並列に記憶するようにしておけば、大容量化及び高速化を図ることができる。さらに、病院と病院以外の検査機関等とでデータを共有しつつ並列処理を実行すれば、より多くの患者に対する検査及び診断の効率化を図ることができる。

【 0 0 4 1 】

上述のように、第 1 の実施形態によるバイオマーカの検出方法はハイスループット技術を用いることが好ましく、ハイスループット技術に対応可能な装置を用いた方がよい。人間の遺伝子は 2 万個以上あるが、例えば、複数の因子項目として 2 万個の遺伝子 ($m = 20,000$) に関する発現量を用いる場合、複数の因子項目間の第 1 の相関係数の計算 (S T 3)、複数の因子項目間の第 2 の相関係数の計算 (S T 4)、及び第 1 の相関係数と第 2 の相関係数との差分相関係数の計算 (S T 5) の各ステップにおいて、 $20000 C_2 (= 199,990,000)$ 通りの計算を実行しなければならない。言うまでもなく、これだけの膨大な計算を人間が合理的な時間内で正確に行うことは不可能であり、膨大な計算を正確に実行可能な装置を用いる必要がある。仮に上記の計算を複数の人間が行ったとしても、とてつもない時間を要し（何年にもわたる可能性がある。）、人間が計算している間に患者の病状が悪化するおそれがある。これでは、疾病状態へ遷移する前の過渡的な状態（健康と病気の間の状態）を検出するという目的を達成することができない。したがって、上述のバイオマーカの検出方法を実現するためには、膨大な計算を高速且つ正確に実行可能な検出装置 1 を用いる必要がある。以下に述べる第 3 の実施形態によるバイオマーカの検出方法を実行する装置及び第 4 の実施形態によるバイオマーカの検出方法を実行する装置についても同じことが言える。

【 0 0 4 2 】

（第 3 の実施形態）

次に、図 7 を参照しながら、本発明の第 3 の実施形態によるバイオマーカの検出方法を説明する。図 7 は、第 3 の実施形態におけるバイオマーカの検出方法を示すフローチャートである。

【 0 0 4 3 】

第 3 の実施形態によるバイオマーカの検出方法は、図 7 に示すように、ハイスループットデータを取得するステップ (S T 1 1) と、差次的因子の選出するステップ (S T 1 2) と、クラスター化するステップ (S T 1 3) と、バイオマーカを選出するステップ (S T 1 4) とを含んでいる。

初めに、ハイスループットデータを取得するステップ (S T 1 1) において、被検者 S e から異なる時点でハイスループットデータが得られる。具体的には、まず、被検者 S e から生体サンプルが時系列的に複数回採取される。生体サンプルを採取する回数は、例えば 5 回以上が望ましい。また、各回の間隔は、疾病の状況によって、数日、数週間、数カ月、又は数年に設定して良く、例えば数分又は数時間としても良い。

このようにして得られた複数の生体サンプルのそれぞれから、因子項目としての遺伝子の発現量が、ハイスループットデータとしてハイスループット技術により取得される。

【 0 0 4 4 】

次に、差次的因子を選出するステップ (S T 1 2) において、取得されたハイスループットデータから差次的因子が選出される。すなわち、ハイスループット技術により 1 つの生体サンプルから取得された 2 万個以上の遺伝子データの中から、差次的な遺伝子を選出される。ここで差次的な遺伝子は発現量に著しい変化を示した遺伝子である。その選出は、具体的には以下のように行われる。まず、複数の生体サンプルの中の初回に採取された生体サンプルをコントロールサンプルとする。説明の便宜上、このコントロールサンプルから取得されたデータを $D d m c$ とし、2 回目に採取された生体サンプルの遺伝子のデー

タを $D d m$ と呼ぶこととする。次に、この遺伝子のデータ $D d m$ に対して学生 t 検定を行って、コントロールサンプルのデータ $D d m c$ に比べて発現量に著しい変化を示した遺伝子を選出する。説明の便宜上、この遺伝子を遺伝子 $D d m 1$ と呼ぶ。

【 0 0 4 5 】

なお、第 3 の実施形態では、遺伝子 $D d m 1$ の選出に学生 t 検定を用いたが、他の実施形態では、例えば、マン・ホイットニーの U 検定等の他の検定方法を適用することも可能である。このようなノンパラメトリック手法による検定は、母集団となる $D d m$ が正規分布に従わない場合に、特に有効である。また、学生 t 検定をする場合においても、有意水準の値は、 0.05 、 0.01 等の値を適宜設定することができる。

10

【 0 0 4 6 】

次いで、クラスター化するステップ ($S T 1 3$) において、差次的因子を選出するステップ ($S T 1 2$) において選出された差次的因子 (差次的な遺伝子) が複数のクラスターに分類される。

このステップ ($S T 1 2$) においては、初めに、コントロールサンプルのデータ $D d m c$ に比べて発現量に著しい変化を示した遺伝子 $D d m 1$ に対し、誤発現率 $F D R$ (False Discovery Rate) を用いて、複数の比較 (Multiple Comparisons) を行うことにより、又は複数の学生 t 検定を補正することにより、遺伝子 $D d m 2$ (図示なし) が得られる。次に、2 倍変更方法 (Two fold Change Method) を用いて、遺伝子 $D d m 2$ から、標準偏差 $S D$ が比較的著しく変化する遺伝子が差次的な遺伝子 $D s m$ (図 8) として選出される ($S T 1 2$)。ここで選出された差次的な遺伝子 $D s m$ は、コントロールサンプルのデータ $D d m c$ に比べて著しい差異を示すだけでなく、自分の平均値からも大きく逸脱している。

20

【 0 0 4 7 】

次に、差次的な遺伝子 $D s m$ の間の相関係数が計算される。ここで、説明の便宜上、遺伝子 $g n 1$ 、 $g n 2$ 、 $g n 3$ 、 \dots 、 $g n p$ が差次的な遺伝子 $D s m$ として選出されたと仮定する。また、遺伝子 $g n 1$ のデータ (発現量) は、生体サンプルの採取の時点 $t 1$ 、 $t 2$ 、 $t 3$ 、 \dots 、 $t p$ において、順に $g n 1 1$ 、 $g n 1 2$ 、 $g n 1 3$ 、 \dots 、 $g n 1 p$ であったと仮定すると、 p 個の要素を有する集合 $C L 1 m : \{ g n 1 1, g n 1 2, g n 1 3, \dots, g n 1 p \}$ が得られる。同様に、遺伝子 $g n 2$ の発現量が、生体サンプルの採取の時点 $t 1$ 、 $t 2$ 、 $t 3$ 、 \dots 、 $t p$ において、順に $g n 2 1$ 、 $g n 2 2$ 、 $g n 2 3$ 、 \dots 、 $g n 2 p$ であったと仮定すると、 p 個の要素を有する集合 $C L 2 m : \{ g n 2 1, g n 2 2, g n 2 3, \dots, g n 2 p \}$ が得られる。他の遺伝子 $g n 3$ 、 \dots 、 $g n p$ についても同様な集合 $C L 3 m$ 、 \dots 、 $C L p m$ が得られる。

30

【 0 0 4 8 】

次いで、すべての遺伝子の集合 $C L 1 m$ 、 $C L 2 m$ 、 $C L 3 m$ 、 \dots 、 $C L p m$ のうちのあらゆる一対の集合の間で相関係数が計算される。この相関係数はピアソン相関係数 ($P C C$) であっても良い。計算された相関係数 $P C C$ のうち、閾値を超えるものが選出され、選出された $P C C$ に基づき、例えばネットワークが構築される。例えば、集合 $C L 1 m$ 及び $C L 2 m$ の間の相関係数 $P C C$ が閾値を超えた場合、遺伝子 $g n 1$ に対応するノード $G n 1$ と、遺伝子 $g n 2$ に対応するノード $G n 2$ とが設定され、これらのノード $G n 1$ 及び $G n 2$ が枝 B でリンクされる。同様な操作が、閾値を超えたすべての $P C C$ に対して行われ、図 8 に示すネットワーク $N m$ が構築される。

40

【 0 0 4 9 】

この後、ネットワーク $N m$ においてノードがクラスター化される。すなわち、複数のノードがグループ (クラスター) に分類される。ここでいうクラスター化とは、複数の生体分子を、互いに相関の高いグループにそれぞれ分類する処理であり、生体分子が分類されたそれぞれのグループをクラスターと称する。すなわち、互いに相関の高い生体分子を 1 つのクラスターにするように、上述の差次的な遺伝子 $D s m$ を、 n 個のクラスターに分類する。図 8 を参照すると、点線で画定された 5 つのクラスターと、各クラスターに含まれ

50

る2つ又は3つのノードとが示されている。ただし、クラスターの数や、クラスターに含まれるノードの数は、図8の例に限定されることなく、適宜決定しても良い。例えばクラスターの数全部で3個以上であって良く、一つのクラスターには5個以上のノードが含まれても良い。

【0050】

次いで、バイオマーカーの候補を選出するステップ(ST14)において、クラスター化するステップ(ST13)で得られた複数のクラスターの中から、クラスターの内側の因子項目の間の相関の増大、クラスターの内側の因子項目の標準偏差の増大、及びクラスターの内側の因子項目と外側の因子項目との間の相関の低減が最も著しいクラスターがバイオマーカーとして選出される。

具体的には、各クラスターについて、クラスターの内側のノード間の相関係数の平均値 PCC_{in} (以下、内部 PCC 平均値と呼ぶ) が計算され、クラスターの内側のノードと外側のノードとの間の相関係数の平均値 PCC_{out} (以下、内外 PCC 平均値と呼ぶ) が計算され、内側のノードの標準偏差 SD_{in} が計算される。図8を参照すると、ノード G_{n1} 、 G_{n2} 、及び G_{n3} を含むクラスター C が形成されている。このクラスター C に着目すると、クラスター C の内側のノード G_{n1} 及び G_{n2} の間の相関係数 (すなわち、遺伝子 g_{n1} 及び g_{n2} の間の相関係数 PCC) と、ノード G_{n2} 及び G_{n3} の間の相関係数 (遺伝子 g_{n2} 及び g_{n3} の間の相関係数 PCC) と、ノード G_{n3} 及び G_{n1} の間の相関係数 (遺伝子 g_{n3} 及び g_{n1} の間の相関係数 PCC) とが平均化され、内部 PCC 平均値 PCC_{in} が得られる。

【0051】

続けて、ノード G_{n1} と、これに直接に接続するもののクラスター C の外側にあるノード G_{n4} との間の相関係数 (遺伝子 g_{n1} 及び g_{n4} の間の相関係数 PCC)、ノード G_{n2} と、これに直接に接続するもののクラスター C の外側にあるノード G_{n5} との間の相関係数 (遺伝子 g_{n2} 及び g_{n5} の間の相関係数 PCC) と、ノード G_{n3} と、これに直接に接続するもののクラスター C の外側にあるノード G_{n6} 及び G_{n7} のそれぞれとの間の相関係数 (遺伝子 g_{n3} 及び g_{n6} の間の相関係数 PCC 、遺伝子 g_{n3} 及び g_{n7} の間の相関係数 PCC) といった、合計4つの相関係数が平均化され、内外 PCC 平均値 PCC_{out} が得られる。

さらに、クラスター C の内部のノード G_{n1} 、 G_{n2} 、及び G_{n3} のそれぞれに対応する遺伝子 g_{n1} 、 g_{n2} 、及び g_{n3} のデータ (発現量) の標準偏差 SD_{in} が計算される。

【0052】

以上のようにして求めた内部 PCC 平均値 PCC_{in} 、内外 PCC 平均値 PCC_{out} 、及び標準偏差 SD_{in} から、以下の式(2)に従って、クラスター C の指標 I_m が計算される。

$$I_m = SD_{in} \times PCC_{in} / PCC_{out} \quad \dots \quad \text{式(2)}$$

続けて、同様の計算がすべてのクラスターに対して行われ、最も大きな指標 I_m を提供したクラスター (又は、そのクラスターに属するノードに対応する遺伝子群) がバイオマーカーとして選択される。

なお、同一のノード (群) に対し、クラスターの数や、クラスターに含まれるノードの数を変えて、同様の計算を複数回繰り返し、更に大きな指標 I_m を提供するクラスターを見出しても良い。

また、各クラスターの指標 I_m に対して有意性分析を行なってクラスターがバイオマーカーであるか否かを判定しても良い。

なお、クラスター化を行わないで、それぞれの遺伝子について指標を計算することも考えられ、上述のステップ $ST12$ において差次的な遺伝子 D_{sm} を選出する際に計算した標準偏差 SD を用いて、指標を計算することができる。具体的には、上記の式(1)の分子の sED を、この標準偏差 SD に置き換えることにより得られる下記の式(3)に従って、指標 I_t を求めても良い。

10

20

30

40

50

$$I t = S D \times s P C C i n / s P C C o u t \quad \dots \quad \text{式(3)}$$

すべてのノード（差次的な遺伝子 $D s m$ ）について計算された複数の指標 $I t$ が大きい順にソートされ、ソートされた指標 $I t$ から大きい順に所定の数の指標 $I t$ が選択される。これにより、選択された所定の数の指標 $I t$ に対応する遺伝子がバイオマーカーとして検出されたこととなる。

【0053】

以上のように第3の実施形態によるバイオマーカーの検出方法においては、被検者 $S e$ から時系列的に採取された複数の生体サンプルから遺伝子のデータ（発現量）が取得され、その中から差次的な遺伝子 $D s m$ が選出される。差次的な遺伝子 $D s m$ について、相関係数 $P C C$ が得られ、相関係数 $P C C$ に基づいてネットワーク $N m$ が構築される。ネットワーク $N m$ を構成するノードがクラスターへとグループ分けされ、各クラスターについて指標 $I m$ が計算され、その値が最大となるクラスターに含まれるノードに対応する遺伝子がバイオマーカーとして選択される。

また、クラスター化を行わないで、それぞれの遺伝子について指標を計算する場合は、差次的な遺伝子 $D s m$ を選出する際（ステップ $S T 1 2$ ）に計算した標準偏差 $S D$ を用いる上記の式（3）により指標 $I t$ を計算できる。この標準偏差 $S D$ は、注目している遺伝子の発現量が多くなった、あるいは急激に多くなったことを表現できる変数であるため、式（1）の $s E D$ の代わりとなることができる。

【0054】

第3の実施形態によれば、差次的な因子（発現量に著しい変化を示した遺伝子 $D s m$ ）を選出し、これらの間の相関係数を計算し、クラスター化し、クラスターについて指標 $I m$ が最大となるクラスターがバイオマーカーとして選択される、又はクラスター化を行わないで、それぞれの遺伝子について指標を計算する場合は、差次的な因子を選出する際に計算した標準偏差 $S D$ を用いて指標 $I t$ （式（3））を計算し、その値が大きい順に1又は2以上の所定の数の指標 $I t$ が選択されるため、互いに強く相関する複数のノードを求めることができる。過渡的な状態においては、一般に、揺らぎが増大するとともに、その揺らぎが強く相関する変数が存在することが知られている。したがって、第3の実施形態によるバイオマーカーの検出方法によりバイオマーカーが検出されたことをもって、被検者 $S e$ が過渡的な状態にあることが分かる。

【0055】

（第4の実施形態）

次に、図9を参照しながら、本発明の第4の実施形態によるバイオマーカーの検出方法を説明する。第4の実施形態のバイオマーカーの検出方法は、被検者 $S e$ の生体サンプルを採取する回数の中で、第1の実施形態によるバイオマーカーの検出方法と異なる。具体的には、第1の実施形態のバイオマーカーの検出方法においては、被検者 $S e$ の生体サンプルの採取回数は1回であったが、第4の実施形態のバイオマーカーの検出方法では、時系列的に複数回採取される。以下、相違点を中心に第4の実施形態のバイオマーカーの検出方法を説明する。

【0056】

図9を参照すると、異なる時点 $t 1$ 、 $t 2$ 、 $t 3$ で被検者 $S e$ から採取された生体サンプルから取得された被検者データセット $D e m$ が得られている。生体サンプルを採取する回数は、説明の便宜上、3回としたが、例えば5回以上が望ましい。また、各回の間隔は、疾病の状況によって、数日、数週間、数カ月、又は数年に設定して良く、例えば数分又は数時間としても良い。

被検者データセット $D e m$ は、第1の実施形態と同様に予め用意された参照データセット $D r$ に追加され、検査対象データセット $D c m$ が作成される。本例では、検査対象データセット $D c m$ において各遺伝子の集合に属するデータの数は $n + 3$ である。

【0057】

次いで、第1の実施形態における第1の相関係数を計算するステップ（ $S T 3$ ）及び第2の相関係数を計算するステップ（ $S T 4$ ）と同様にして、参照データセット $D r$ 内のデ

ータ間の相関係数 $PCCr$ と、検査対象データセット Dcm 内のデータ間の相関係数 PCc とが計算され、これらの中で差分相関係数 PC が計算される。次いで、第1の実施形態における差分相関係数を計算するステップ (ST5) と同様にして、 PC のうち、閾値を超えた PC が抽出され、図4に示す差分相関係数ネットワーク Nrc が構築される。

【0058】

差分相関係数ネットワーク Nrc に関して、第1の実施形態における、指標を求めるステップ (ST6) と同様に、着目するノードと1次のノードとの間の相関係数の平均値 $sPCCin$ と、着目するノードに関する1次のノードと2次のノードとの間の相関係数の平均値 $sPCCout$ が求められる。

図9を参照すると、グラフ GRm には、遺伝子 $g1$ についての発現量 - 確率曲線が破線で示されている。このグラフにおいては、参照データセット Dr 中の遺伝子 $g1$ についての発現量 - 確率曲線 Cr と、被検者データセット Dem 中の遺伝子 $g1$ についての発現量 - 確率曲線 Ce とが示されている。ここで、平均値 r は、発現量 - 確率曲線 Cr の平均値を表し、平均値 e は、発現量 - 確率曲線 Ce の平均値を表している。このグラフから、平均値 r と e の差 $sED(|r - e|)$ が求められる。

【0059】

次に、上述の式(1)に従って、上記の平均値 $sPCCin$ と、平均値 $sPCCout$ と、差 sED とから、ノード $G1$ (すなわち、遺伝子 $g1$) についての指標 Is が求められる。

次いで、すべてのノード $G2$ 、 $G3$ 、 \dots 、 $G10$ (すなわち、対応する遺伝子 $g2$ 、 $g3$ 、 \dots 、 $g10$) について、同様に指標 Is が計算される。

【0060】

次に、図4に模式的に示すように、すべてのノードについて計算された複数の指標 Is が大きい順にソートされる。ソートされた指標 Is から大きい順に所定の数の指標 Is が選択される。選択された所定の数の指標 Is に対応する遺伝子がバイオマーカーとして検出されたこととなる。選択される指標 Is の数は、閾値を定めることにより決定し、閾値を超える指標 Is が選択されるようにしても良い。

【0061】

第4の実施形態によるバイオマーカーの検出方法によっても、参照データセット Dr の因子項目間の相関係数 $PCCr$ と、検査対象データセット Dcm の因子項目間の相関係数 PCc との差分相関係数 PC が求められ、これに基づいた指標からバイオマーカーが検出され得るため、第1の実施形態によるバイオマーカーの検出方法と同様の効果が発揮される。

また、上記の実施形態によるバイオマーカーの検出方法又は検出装置により探索されたバイオマーカーを使えば、ある疾病についての判定(病気になる直前の状態か、あるいは病気から回復する直前の状態かの判定)を行うことができる。判定の対象となる人に対して検査を行い、採取した生体サンプルからバイオマーカーを計算で求め、求めたバイオマーカーとすでに探索されたバイオマーカーとを比較する。例えば、すでに探索され、所定の疾病に対応したバイオマーカーの遺伝子群(上位の遺伝子群、あるいはあらかじめ特定しておいた遺伝子群であって、遺伝子群に含まれる遺伝子は1つでも良い)に、求めたバイオマーカーに含まれているか否かを判定する。

この判定は、たとえば第2の実施形態によるバイオマーカー検出装置等により自動で判定しても良いし、医者が判定しても良い。

また、判定結果により疾病に対応するバイオマーカーあるいは遺伝子であると判定した場合、当該疾病に係る情報を提示しても良い。

これにより、その人が、所定の病気の直前の状態にあるのか、あるいは所定の病気から回復する直前の状態にあるのか判定できる。

【0062】

以上、本発明によるバイオマーカーの検出方法、バイオマーカーの検出装置、及びバイ

10

20

30

40

50

オマーカーの検出プログラムの幾つかの実施形態を説明したが、本発明は上述の幾つかの実施形態に限定されることなく、特許請求の範囲内で種々に変形又は変更が可能である。

たとえば、第1及び第4の実施形態では、参照データセット D_r は、被検者 S_e の生体サンプルを採取する前に用意されたが、他の実施形態においては、被検者 S_e の生体サンプルの採取後、又は、この生体サンプルから遺伝子 g_1 、 g_2 、 g_3 、 \dots 、 g_m に関する被検者データセット D_e が取得された後に用意されてもかまわない。また、参照データセット D_r を予め作成し、データベースに登録しておいても良い。このようにすれば、必要に応じてデータベースから例えばインタフェイス部15を通して記憶部11にダウンロードすることによって、参照データセット D_r を用意できる。

10

【0063】

また、第1及び第4の実施形態において、指標 I_s は式(1)に従って計算されたが、この式の分母($sPCCout$)を1としても(すなわち、分子のみに基づいて指標 I_s を計算しても)、バイオマーカーを選択できることが発明者らの検討により明らかとなっている。

【0064】

さらに、第1及び第4の実施形態においては、見掛け上は健康が良好であることを前提に、被検者 S_e が過渡的な健康状態にあるかどうか調べられる。他の実施形態では、それとは逆に、被検者 S_e が特定の病気に罹っていることを前提としても良い。この場合には、被検者 S_e と同じ病気に罹っている患者が生体データ提供者 S_1 、 S_2 、 \dots 、 S_n になることができる。このような場合において、被検者 S_e からバイオマーカーが検出されたときは、被検者 S_e の健康は、疾病状態(図1の点d又はc)から良好な状態(同図の点a)に向かう過渡的な状態にあり、被検者 S_e は快復に向かっていることが推測される。なお、特定の病気とは、ある病気に限定されることはなく、どんな病気であっても良い。

20

また、被検者 S_e が特定の病気に罹り、その病気に対して特定の治療を受けている場合には、同じ治療を受けている患者が生体データ提供者 S_1 、 S_2 、 \dots 、 S_n になることもできる。この場合において、被検者 S_e からバイオマーカーが検出されたときは、その治療が被検者 S_e に対し効果を発揮しており、疾病状態(図1の点d又はc)から良好な状態(同図の点a)に向かう過渡的な状態にあり、被検者 S_e は快復に向かっていることが推測される。

30

【0065】

また、第3の実施形態においても、健康状態が良好な被検者 S_e についてバイオマーカーを検出できた場合には、その被検者 S_e の健康は、良好な状態(同図の点a)から疾病状態(図1の点d又はc)へ向かう過渡的な状態(図1の点b)にあることが分かる。これとは逆に、被検者 S_e が疾病状態のときにバイオマーカーが検出されれば、その被検者 S_e の健康は過渡的な状態にあり、被検者 S_e は快復に向かっていることが推測される。

【0066】

このように被検者 S_e が疾病状態にある場合において、患者自身が病状の改善を感じることができないとき、または従来の検査によっては病状の改善が認められないときであっても、本発明の実施形態によるバイオマーカーの検出方法によれば、薬が効いているかどうかを確認することも可能である。そのため、治療の継続の可否について正しく判断することが可能となる。

40

【0067】

また、従来の検査によっては病状の改善が認められるまでに長い期間を要する場合がある。この場合において、長い期間が経過した後に薬効がなかったことが判明したときには、その期間に病状が悪化してしまうことにもなる。本発明の実施形態によるバイオマーカーの検出方法を用いれば、状態遷移の予兆を早期に把握できるため、薬の使用開始後の短い期間で薬効の有無を判断できる。そのため、薬効がない場合に、無駄な治療をせずに済むし、早期に他の治療方法に変更することができ、病気の進行の遅延や病状の改善につな

50

げることとも可能となる。また、高額な医薬品を長期に亘って無益に使用するのを回避することが可能となり、医療費を抑制し得る。

【 0 0 6 8 】

また、第 4 の実施形態によるバイオマーカーの検出方法は、第 2 の実施形態によるバイオマーカーの検出装置 1 において実施しても良い。この場合、第 4 の実施形態によるバイオマーカーの検出方法を検出装置 1 に実行させる、バイオマーカーの検出プログラムが記憶部 1 1 に記録されていても良い。また、この検出プログラムは、非一時的又は有形のコンピュータ可読記憶媒体に記録されても良い。

【 0 0 6 9 】

また、第 4 の実施形態においては、被検者 S_e の生体サンプルから時系列的に取得された複数のデータから被検者データセット D_{em} を作成し、これを参照データセット D_r に追加して検査対象データセット D_{cm} を作成したが、他の実施形態においては、異なる時点でのデータを単独で使用することもできる。すなわち、生体サンプルを異なる時点で採取するごとに、第 1 の実施形態によるバイオマーカーの検出方法を実施しても良い。より具体的には、時点 t_1 において被検者 S_e から生体サンプルを採取した際に第 1 の実施形態によるバイオマーカーの検出方法を実施し、これを時点 t_2 、 t_3 、 \dots のそれぞれにおいて繰り返すようにしても良い。このような検出方法もまた、検出プログラムにより上述の検出装置 1 で実施することができ、この検出プログラムは、非一時的又は有形のコンピュータ可読記憶媒体に記録されても良い。

【 0 0 7 0 】

また、第 1、第 3、及び第 4 の実施形態において、差分相関係数ネットワーク N_{rc} 又はネットワーク N_m が構築される例を説明したが、ネットワークを構築しなくてもノード間の相関係数を計算することは可能であり、したがって、ネットワークは必ずしも構築しなくても良い。

また、第 1 及び第 4 の実施形態において、差分相関係数ネットワーク N_{rc} について式 (1) を用いて指標 I_s を計算する例を説明したが、差分相関ネットワーク N_{rc} のノードをクラスター化し、式 (2) に従って、指標 I_m を計算し、バイオマーカーを選択することも可能である。

すなわち、第 1 及び第 4 の実施形態によるバイオマーカー検出方法において、差分相関係数ネットワーク N_{rc} について式 (1) を用いて指標 I_s を計算する場合と、差分相関係数ネットワーク N_{rc} のノードをクラスター化し、式 (2) に従って、指標 I_m を計算する場合と、第 3 の実施形態によるバイオマーカー検出方法において、差次的な遺伝子 D_{sm} についての相関係数 P_{CC} に基づいて構築されたネットワーク N_m に対し、式 (1) を用いて指標 I_s を計算する場合と、このネットワーク N_m のノードをクラスター化し、式 (2) に従って、指標 I_m を計算する場合との合計 4 通りのバイオマーカー検出方法がある。

さらに、第 3 の実施形態においてハイスループット技術により取得されたハイスループットデータから差次的因子を選択したが、第 1 及び第 4 の実施形態においても、差次的因子を選択しても良い。また、上述の sED を用い、たとえば所定の値以上の sED を有する因子を差次的因子として選択しても良い。

【 0 0 7 1 】

また、上記の実施形態においては、説明の便宜上、因子項目として遺伝子 (の発現量) を例示したが、遺伝子に関わらず、血液検査における所定の検査項目 (タンパク質やコレステロール、血糖値などの測定項目、代謝物に関する測定項目など) の数値を用いることができる。また、生体サンプルを採取し、生体サンプルから因子項目を抽出する場合には、生体サンプルとしては、血液に限らず、唾液や汗、又は尿や便などの排泄物などであっても良く、生体組織 (例えば肝臓障害の場合、肝臓の組織) であっても良い。さらに、心電図や脳波、コンピュータ断層撮影 (CT) や (核) 磁気共鳴映像 (MRI)、PET (Positron Emission Tomography) 画像などの信号を数値化することにより得られた数値などを用いることができる。さらにまた、声や心音のような体内から発せられる音を測定し

、測定値を数値化することにより得られた数値などを用いることも可能である。なお、上の説明で因子項目という語を用いたが、これは、上述のネットワークのノードとなり得る項目をいう。

【0072】

本発明によるバイオマーカーの検出方法は、総合病院や大学病院などで実施することができるが、病院外の検査機関、個人向けの血液診断サービスや遺伝子診断サービスを提供している民間の企業等によっても実施されても良い。このような検査機関及び企業等において、本発明によるバイオマーカーの検出装置や検出プログラムを使用できることはもちろんである。

また、本発明によるバイオマーカーの検出方法、検出装置、及び検出プログラムは、人間だけでなく、動物など生物一般に対しても適用することができる。

10

【0073】

なお、従来から疾病状態を診断するために用いられる指標としてバイオマーカーがある。バイオマーカーは良好状態（基準値内）と病気状態（基準値外）を判別したり、病気状態の変化（改善又は悪化）を調べたりするのに用いられる。これに対し、本発明によるバイオマーカーの検出方法では、良好な状態から前記疾病状態への遷移の予兆（過渡的な状態）を検知することができる。すなわち、本発明によるバイオマーカーの検出方法には、病気に罹るのを早期に回避することが可能となるという利点がある。

【0074】

また、従来のバイオマーカーは、生体から採取された血清や尿などの体液や組織に含まれる化学物質であり、分子式又は特性で特定することも可能である。一方、本発明に係るバイオマーカーは、被検者や疾病によって異なり、検出されて初めてその存在が明らかになる。そのため、構成（分子式等）や特性により直接に特定することはほぼ不可能であり、およそ現実的でない。よって、本発明に係るバイオマーカーは従来のバイオマーカーと異なる。

20

【0075】

言い換えると、本発明の他の実施形態によるバイオマーカーは、生体に関する測定により得られたデータに基づいて、生体の状態の指標となるバイオマーカーを検出する、バイオマーカー検出方法であって、複数の参照生体のそれぞれから得られたデータに基づく参照データセットを用意し、対象生体から得られた対象生体データを参照データセットに追加して対象データセットを生成し、参照データセットの中の複数の因子項目の間で第1の相関係数を求め、対象データセットの中の複数の因子項目の間で第2の相関係数を求め、第1の相関係数と第2の相関係数との差分である差分相関係数を求め、複数の因子項目のそれぞれについて差分相関係数に基づく指標を求め、指標に基づいてバイオマーカーを選択する工程を含む、バイオマーカー検出方法により検出されるバイオマーカーである。

30

【0076】

なお、上記の実施形態によるバイオマーカーの検出方法及びバイオマーカーの検出プログラムは、被検者が過渡的な状態にあることを示す（疾病状態の予兆を示す）指標を具体的に見出すものであり、従来から知られていた日常的な生体データ処理をコンピュータで単に履行するものではない。

40

また、上述してきたバイオマーカーは動的ネットワークバイオマーカー（Dynamic Network Biomarker：DNB）と呼ぶことがある。

【符号の説明】

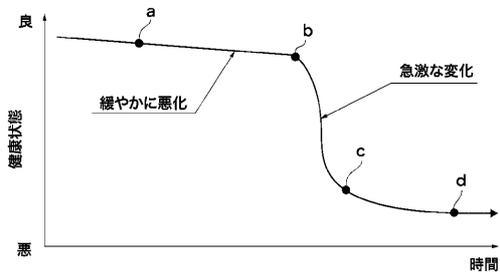
【0077】

- D データ
- D_c、D_{cm} 検査対象データセット
- D_e、D_{em} 被検者データセット
- D_r 参照データセット
- g₁、g₂、・・・、g_m 遺伝子
- G₁、G₂、G₃、G₄、・・・、G₁₀ ノード

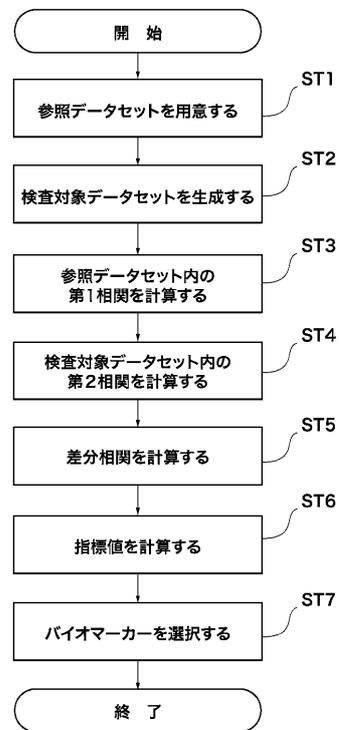
50

- Nrc、Nm 差分相関係数ネットワーク
- Se 被検者
- S1、S2、・・・、Sn 生体データ提供者
- 1 検出装置
- 10 制御部
- 11 記憶部
- 11a 検出プログラム
- 12 記憶部
- 13 入力部
- 13a コンピュータ可読記憶媒体
- 14 出力部
- 15 インタフェイス部

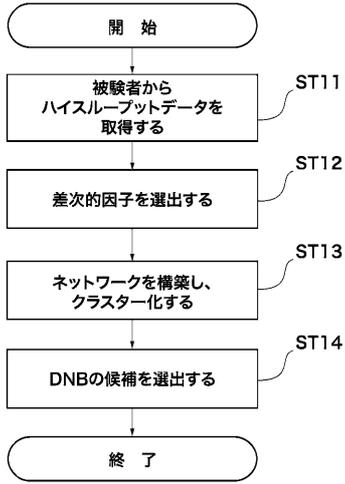
【図1】



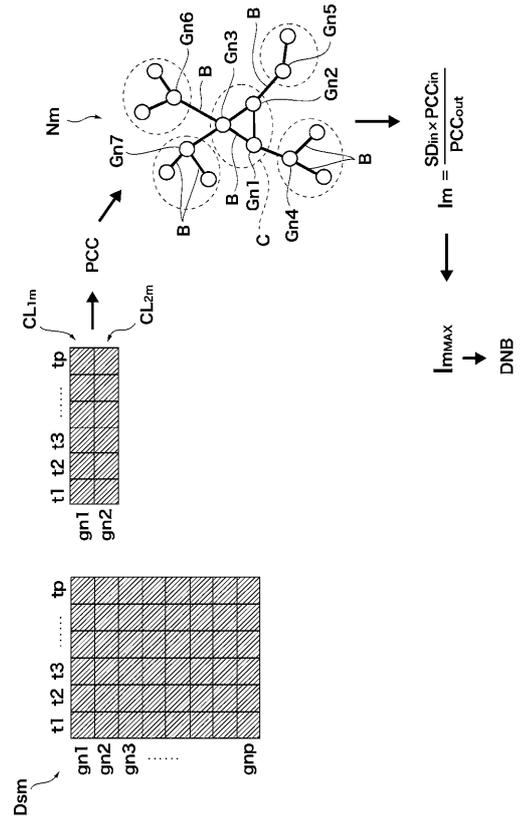
【図2】



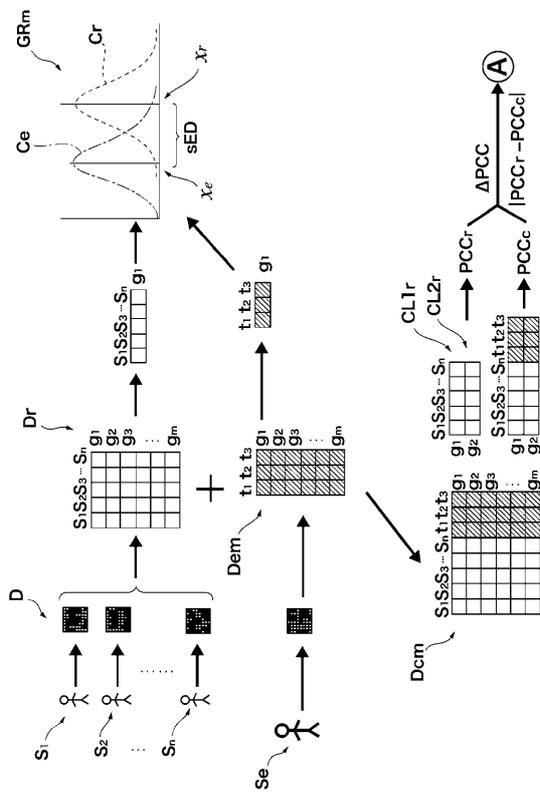
【 図 7 】



【 図 8 】



【 図 9 】



フロントページの続き

(72)発明者 陳 洛南

日本国東京都目黒区駒場4丁目6番1号 国立大学法人東京大学生産技術研究所内

(72)発明者 合原 一幸

日本国東京都目黒区駒場4丁目6番1号 国立大学法人東京大学生産技術研究所内

(72)発明者 劉 小平

日本国東京都目黒区駒場4丁目6番1号 国立大学法人東京大学生産技術研究所内

審査官 松野 広一

(56)参考文献 特開2014-064515(JP, A)

米国特許出願公開第2005/0142573(US, A1)

特開2005-323573(JP, A)

(58)調査した分野(Int.Cl., DB名)

G16B 5/00-99/00

C12M 1/34

G01N 33/48

C12Q 1/68

PubMed