

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4302466号
(P4302466)

(45) 発行日 平成21年7月29日(2009.7.29)

(24) 登録日 平成21年5月1日(2009.5.1)

(51) Int.Cl.		F 1			
C 1 2 N	15/09	(2006.01)	C 1 2 N	15/00	F
G 0 6 F	17/30	(2006.01)	G 0 6 F	17/30	1 7 0 F
G 0 6 F	19/00	(2006.01)	G 0 6 F	19/00	6 0 0

請求項の数 16 (全 32 頁)

(21) 出願番号	特願2003-307587 (P2003-307587)	(73) 特許権者	503360115 独立行政法人科学技術振興機構 埼玉県川口市本町4丁目1番8号
(22) 出願日	平成15年8月29日(2003.8.29)	(74) 代理人	110000338 特許業務法人原謙三国際特許事務所
(65) 公開番号	特開2005-73569 (P2005-73569A)	(74) 代理人	100080034 弁理士 原 謙三
(43) 公開日	平成17年3月24日(2005.3.24)	(74) 代理人	100113701 弁理士 木島 隆一
審査請求日	平成18年4月18日(2006.4.18)	(74) 代理人	100116241 弁理士 金子 一郎
		(72) 発明者	矢野 健太郎 千葉県木更津市東太田4-13-21-2 02

最終頁に続く

(54) 【発明の名称】 発現プロファイル解析システム、発現プロファイル解析方法、発現プロファイル解析プログラム、およびそのプログラムを記録した記録媒体

(57) 【特許請求の範囲】

【請求項1】

遺伝子および/またはタンパク質の発現プロファイルデータの対数変換値を解析する発現プロファイル解析システムにおいて、

上記発現プロファイルデータを対数変換する変換手段と、

上記変換手段によって得られた対数変換値を対応分析によって解析する解析手段とを備え、

上記変換手段は、対数変換の指標として、 $\arctan(1/\text{ratio})$

(ここで、ratioは、任意の表現型での遺伝子またはタンパク質の発現量と、比較対照となる表現型での遺伝子またはタンパク質の発現量の比である。)

を用いることを特徴とする発現プロファイル解析システム。

【請求項2】

上記解析手段は、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第1補足データと、当該第1補足データと反対の発現様式のデータである第2補足データと、いずれの表現型においても等発現するデータである第3補足データとを用いて対応分析を行うことを特徴とする請求項1に記載の発現プロファイル解析システム。

【請求項3】

上記解析手段は、上記第1補足データの対応分析結果と、上記第2補足データの対応分析結果とを通る直線を算出することを特徴とする請求項2に記載の発現プロファイル解析

10

20

システム。

【請求項 4】

上記解析手段は、上記直線から、所定の距離の範囲内にある第 1 有意領域を設定することを特徴とする請求項 3 に記載の発現プロファイル解析システム。

【請求項 5】

上記解析手段は、上記第 3 補足データの対応分析結果から、所定の距離の範囲内にある第 2 有意領域を設定することを特徴とする請求項 2, 3, または 4 に記載の発現プロファイル解析システム。

【請求項 6】

上記解析手段は、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を制御することを特徴とする請求項 1 ~ 5 のいずれか 1 項に記載の発現プロファイル解析システム。

10

【請求項 7】

上記解析手段は、対応分析の解析結果を回転可能な画像データに変換することを特徴とする請求項 1 ~ 6 のいずれか 1 項に記載の遺伝子発現プロファイル解析システム。

【請求項 8】

上記発現プロファイルデータは、マイクロアレイ、マクロアレイ、ディファレンシャルディスプレイの少なくともいずれかによって得られたものであることを特徴とする請求項 1 ~ 5 のいずれか 1 項に記載の発現プロファイル解析システム。

【請求項 9】

遺伝子および/またはタンパク質の発現プロファイルデータの対数変換値を解析する発現プロファイル解析システムの発現プロファイル解析方法において、

上記発現プロファイル解析システムに備えられるコンピュータが、

上記発現プロファイルデータを、 $\arctan(1/\text{ratio})$ を用いて対数変換する変換ステップと、

20

(ここで、ratioは、任意の表現型での遺伝子またはタンパク質の発現量と、比較対照となる表現型での遺伝子またはタンパク質の発現量の比である。)

上記変換ステップによって得られた変換値を対応分析する解析ステップとを実行することを特徴とする発現プロファイル解析方法。

【請求項 10】

上記解析ステップは、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第 1 補足データと、当該第 1 補足データと反対の発現様式のデータである第 2 補足データと、いずれの表現型においても等発現するデータである第 3 補足データとを用いて対応分析を行うことを特徴とする請求項 9 に記載の発現プロファイル解析方法。

30

【請求項 11】

上記解析ステップは、上記第 1 補足データの対応分析結果と、上記第 2 補足データの対応分析結果とを通る直線を算出することを特徴とする請求項 10 に記載の発現プロファイル解析方法。

【請求項 12】

上記解析ステップは、上記直線から、所定の距離の範囲内にある第 1 有意領域を設定することを特徴とする請求項 11 に記載の発現プロファイル解析方法。

40

【請求項 13】

上記解析ステップは、上記第 3 補足データの対応分析結果から、所定の距離の範囲内にある第 2 有意領域を設定することを特徴とする請求項 10, 11, または 12 に記載の発現プロファイル解析方法。

【請求項 14】

上記解析ステップは、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を制御することを特徴とする請求項 9 ~ 13 のいずれか 1 項に記載の発現プロファイル解析方法。

50

【請求項 15】

請求項 1 ~ 8 のいずれか 1 項に記載の発現プロファイル解析システムを動作させるための発現プロファイル解析プログラムであって、コンピュータを上記変換手段および/または解析手段として機能させるための発現プロファイル解析プログラム。

【請求項 16】

請求項 15 に記載の発現プロファイル解析プログラムが記録されたコンピュータ読取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、新規遺伝子発現解析システムおよび遺伝子発現解析方法に関するものである。

【背景技術】

【0002】

ゲノム解析研究の進展により、機能未知の新規遺伝子が大量に同定されている。その機能未知遺伝子の機能を解明するためには、その機能を示唆する情報を得ることが必要となり、その情報を得るためには、遺伝子の発現パターンが重要な役割を果たす。

【0003】

そこで、近年、DNA マイクロアレイや DNA チップ等によって、疾患患者・病態モデル動物の組織・培養細胞内などから取得した数万もの大量の遺伝子の発現が、網羅的に解析されている。マイクロアレイによる遺伝子解析では、発現パターンの特徴から、アレイ上の全遺伝子を網羅的に分類している。この解析には、遺伝子発現プロファイル解析が頻繁に用いられる。

【0004】

一般に、 n 個の遺伝子から構成されたマイクロアレイを使用して、 k 回の独立した実験条件から得られたシグナル強度のデータは、各遺伝子について、 k 次の特徴ベクトルを与える。そして、遺伝子は、これらの特徴ベクトルによって、特徴空間上に座標を指定された n 個の点の集合であるとみなされる。「発現プロファイル解析」とは、特徴空間上にプロットされた点、すなわち、遺伝子（プロテインアレイの場合はタンパク質）を、判別空間上のいくつかのグループに分類することである。言い換えると、生体内における遺伝子発現の情報を統合し、その情報を比較検討するものである。

【0005】

これにより、例えば、正常な状態（健常人）では発現している遺伝子が、ある疾患の患者では全く発現していない、または発現量が増加または減少しているなど、疾患患者に特異的な遺伝子発現を捉えることによって、疾患に関連する遺伝子を取得できる。

【0006】

このように、遺伝子発現プロファイル解析は、機能未知遺伝子の機能予測のために特に重要なツールとなる。

【0007】

遺伝子発現プロファイル解析において解析対象となるデータは、遺伝子発現比の指標を行列化したものである。例えば、各行に遺伝子群、各列にサンプル群（標的とする表現型）を並べたものであり、この行と列が遺伝子発現プロファイルである。なお、サンプルとは、より具体的には、異なる複数の調査個体や同一個体での Time Course 実験で計測した表現型などを示す。例えば、100 種類の遺伝子の発現量を、50 個体で計測したとき、行列 A の要素 A_{ij} (i 行 j 列の値、 $1 \leq i \leq 100$ 、 $1 \leq j \leq 50$) は i 番目の遺伝子についての j 番目の個体が示す発現量を示す。

【0008】

遺伝子発現プロファイル解析における膨大な量のサンプルから得られた結果の解析には、その結果を効率よく解析し、目的とする遺伝子を迅速に見出すための情報処理技術が必要となる。従来、このような技術として、例えば、クラスタリング解析、主成分分析な

10

20

30

40

50

どの特別なクラスタリング解析、系統的解析が行われている（非特許文献 1、2 など）。

【0009】

遺伝子発現プロファイル解析は、遺伝子発現量（発現比）を対数変換して行われる。具体的には、対数変換は、発現レベルの比（発現比、ratio）を対数変換した指標（例えば、 $\log_2(\text{ratio})$ など）とするものであり、マイクロアレイ実験によって、ある遺伝子の発現レベルをサンプル間で比較する場合に、主に用いられる。この対数変換を行う理由としては、例えば、 $\log_2(\text{ratio})$ 変換であれば、1/4 倍、1/2 倍、1 倍（等発現）、2 倍、4 倍といった発現比を -2, -1, 0, 1, 2 と 1 倍を中心として等尺度へ変換でき、研究者にとって理解しやすいこと、統計解析を行う上で妥当であることなどが挙げられる。しかし、研究機関や研究者によって、この対数の底に 2, e, 10 などを用いるなど統一性がなく、Web 10
上などで公開されたデータの直接比較ができないという学際的な問題がある。

【0010】

また、クラスタリング解析は、多次元の特徴ベクトルに基づいて類似の遺伝子発現プロファイルをもつ遺伝子群やサンプル群を同一のクラスターに分割することができる。そのため、クラスタリング解析では、全サンプルを通してほぼ均一な発現レベルを示すハウスキーピング遺伝子は、同一のクラスターを形成するため、発見が容易である。しかしながら、階層的クラスタリングは、遺伝子の数の増加に伴い計算量が多くなること、また、与えられたデータセットに依存して樹形図のトポロジーが変化しやすい、行列の大きさの増加とともに急激に解析時間が長くなり、計算機の CPU およびメモリが必要であるなどの欠点も有している。 20

【0011】

また、そのようにして得られた膨大な量（万のオーダー）のサンプルや遺伝子のクラスターを視覚的に把握することは困難であるという問題点も有している。そのため、現在、主に、ピアソンの相関係数から大規模クラスターからターゲットとなるクラスターのみを取り出す操作が行われている。しかしながら、得られたクラスターの Viewer も必ずしも研究者にとって分かりやすいものではない（図 7 参照）。

【0012】

図 7 に示した two-dimensional-display と呼ばれる Viewer は、各遺伝子と各サンプルを縦横（もしくは、その逆）に並べたものである。そして、各セルの色やその色の濃淡が、対応するサンプルと遺伝子の発現の強弱を示すように、視覚化されている。 30

【0013】

また、主成分分析は、遺伝子発現プロファイルの数値の大きさを直接的に比較する統計手法であり、より高速な解析を行うことが可能である。しかしながら、主成分分析では、高速な解析を行う結果、調査対象の表現型とは無関係なハウスキーピング遺伝子は、各主軸に対して異なるスコア（座標のようなもの）が出力されてしまうため、散布図にプロットした場合にも、検出が困難である。

【非特許文献 1】わかる! 使える! DNA マイクロアレイデータ解析入門、羊土社、Steen Knudsen (著), 塩島 聡 (翻訳), 辻本 豪三 (翻訳), 松本 治 (翻訳)

【非特許文献 2】必ずデータが出る DNA マイクロアレイ実戦マニュアル 基本原理、チップ作製技術からバイオインフォマティクスまで、羊土社、岡崎 康司 (編集), 林崎 良英 40

【発明の開示】

【発明が解決しようとする課題】

【0014】

このように、従来の解析法には種々の問題点が存在するが、特に、解析時間（処理時間）が長くなる、微量な遺伝子発現比に対する検出力が低い（量的形質の検出力が低い）という問題点が多い。

【0015】

具体的には、遺伝子発現プロファイル解析では、 10^3 を越える膨大な量のデータを処理して解析が行われる。しかしながら、そのような膨大なデータを、通常の計算機を用いて迅速に計算することは困難である。その結果、解析時間が長くなってしまふ。 50

【0016】

また、従来、主に用いられている階層的クラスタリング手法では、計算時間を短縮・簡略化するために、サンプル間の発現比が数倍以上もしくは数倍以下である遺伝子群を恣意的に注視している。これは、発現量が2～3倍などと大きく変化している遺伝子ほど明らかにサンプル間の表現型の差異に影響を及ぼしているであろうという期待に基づいている。

【0017】

ところが、この手法では、発現比が有意に異なっても差異が小さい遺伝子群が解析対象からの排除されてしまう。その結果、例えば、量的形質に關与する遺伝子を検出することは極めて困難である。すなわち、この手法では、検出しようとする表現型が、定性的ではなく定量的である場合、その表現型に關与する遺伝子のうち、極わずかに遺伝子発現量の比が変化した遺伝子を検出することができない。つまり、従来の手法では、標的とする表現型に關与する遺伝子を全て検出しているとはいえない。

10

【0018】

このように、現在の解析的な立場には、極わずかに発現比が変化した遺伝子を、網羅的に発見するという視点が存在しないため、従来の解析手法（対数変換）では、微量な遺伝子発現比に対する検出力が低いという課題自体が存在しない。

【0019】

しかしながら、マイクロアレイなどによって得られる発現プロファイルデータの大部分は、標的とする表現型に關与しないデータである。従来検出されていなかった量的形質に關与する遺伝子の中には、重要な新規遺伝子が含まれている可能性が高い。それゆえ、量的形質に關連する遺伝子の発見に効果的な新規大規模解析ツールの開発が、必要不可欠である。

20

【0020】

そこで、本願発明は、上記従来の課題に鑑みてなされたものであり、その目的は、膨大な量の発現プロファイルデータを、通常のコンピュータを用いた場合であっても迅速に解析するとともに、従来は、解析対象から排除されていた比較的発現比の小さい遺伝子またはタンパク質も検出可能な、発現プロファイル解析システムおよび解析方法を提供することにある。

【課題を解決するための手段】

30

【0021】

本発明にかかる発現プロファイル解析システムは、上記の課題を解決するために、遺伝子および/またはタンパク質の発現プロファイルデータの対数変換値を解析する発現プロファイル解析システムにおいて、上記発現プロファイルデータを対数変換する変換手段と、上記変換手段によって得られた対数変換値を対応分析によって解析する解析手段とを備え、上記変換手段は、対数変換の指標として、 $\arctan(1/\text{ratio})$ （ここで、ratioは、任意の表現型での遺伝子またはタンパク質の発現量と、比較対照となる表現型での遺伝子またはタンパク質の発現量の比である。）を用いることを特徴としている。

【0022】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段は、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第1補足データと、当該第1補足データと反対の発現様式のデータである第2補足データと、いずれの表現型においても等発現するデータである第3補足データとを用いて対応分析を行うことを特徴としている。

40

【0023】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段は、上記第1補足データの対応分析結果と、上記第2補足データの対応分析結果とを通る直線を算出することを特徴としている。

【0024】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段

50

は、上記直線から、所定の距離の範囲内にある第1有意領域を設定することを特徴としている。

【0025】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段は、上記第3補足データの対応分析結果から、所定の距離の範囲内にある第2有意領域を設定することを特徴としている。

【0026】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段は、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を制御することを特徴としている。

10

【0027】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記解析手段は、対応分析の解析結果を回転可能な画像データに変換することを特徴としている。

【0028】

本発明にかかる発現プロファイル解析システムは、上記の構成に加えて、上記発現プロファイルデータは、マイクロアレイ、マクロアレイ、ディファレンシャルディスプレイの少なくともいずれかによって得られたものであることを特徴としている。

【0029】

本発明にかかる発現プロファイル解析方法は、上記の課題を解決するために、遺伝子および/またはタンパク質の発現プロファイルデータの対数変換値を解析する発現プロファイル解析システムの発現プロファイル解析方法において、変換手段が、上記発現プロファイルデータを、 $\arctan(1/\text{ratio})$ を用いて対数変換する変換ステップと、(ここで、ratioは、任意の表現型での遺伝子またはタンパク質の発現量と、比較対照となる表現型での遺伝子またはタンパク質の発現量の比である。)解析手段が、上記変換ステップによって得られた変換値を対応分析する解析ステップとを含むことを特徴としている。

20

【0030】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第1補足データと、当該第1補足データと反対の発現様式のデータである第2補足データと、いずれの表現型においても等発現するデータである第3補足データとを用いて対応分析を行うことを特徴としている。

30

【0031】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記第1補足データの対応分析結果と、上記第2補足データの対応分析結果とを通る直線を算出することを特徴としている。

【0032】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記直線から、所定の距離の範囲内にある第1有意領域を設定することを特徴としている。

【0033】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記第3補足データの対応分析結果から、所定の距離の範囲内にある第2有意領域を設定することを特徴としている。

40

【0034】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を制御することを特徴としている。

【0035】

本発明の発現プロファイル解析プログラムは、上記の課題を解決するために、上記いずれかの発現プロファイル解析システムを動作させるための発現プロファイル解析プログラ

50

ムであって、コンピュータを上記変換手段および/または解析手段として機能させるためのものである。

【0036】

本発明の記録媒体は、上記の課題を解決するために、上記発現プログラム解析プログラムが記録されたコンピュータ読取り可能にしたものである。

【発明の効果】

【0037】

本発明にかかる発現プロファイル解析システムは、以上のように、変換手段により $\arctan(1/\text{ratio})$ を用いて対数変換するので、発現比が小さくても、対数変換値の変化量が、従来の対数変換値($\log(\text{ratio})$)よりも、著しく大きくなる。この変換は、通常のコンピュータなどで、変換可能である。また、解析手段は、対応分析を行うので、互いに関連する項目が近くに集合する。すなわち、表現型に關与する遺伝子またはタンパク質を容易に検出できる。

10

【0038】

それゆえ、膨大な数の発現プロファイルデータを、迅速に解析できると共に、発現比が小さい遺伝子やタンパク質も解析対象としているので、標的とする表現型に關与する遺伝子やタンパク質を網羅的に解析できるという効果を奏する。とりわけ、量的形質に關与する遺伝子またはタンパク質を、網羅的に検出することが可能である。また、発現プロファイルデータを、標的とする表現型に關与するデータと、關与しないデータとに分類できるという効果を奏する。

20

【0039】

本発明にかかる発現プロファイル解析システムは、以上のように、解析手段が、さらに、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第1補足データと、当該第1補足データと反対の発現様式のデータである第2補足データと、いずれの表現型においても等発現するデータである第3補足データとを用いて対応分析を行う。これにより、第1および第2補足データの対応分析結果を表現型に關与する遺伝子またはタンパク質の検出に、第3の補足データの対応分析結果を表現型に關係なく常に一定量存在する遺伝子やタンパク質(例えば、ハウスキーピング遺伝子)の検出に適用できる。それゆえ、表現型に關与するデータの分類の信頼性が向上するという効果を奏する。

30

【0040】

本発明の発現プロファイル解析システムは、以上のように、解析手段が、さらに、上記第1補足データの対応分析結果と、上記第2補足データの対応分析結果とを通る直線を算出する。これにより、この直線上のデータは標的とする表現型に關与するデータとして検出し、この直線から外れたデータは關与しないデータとして分類できるという効果を奏する。

【0041】

本発明の発現プロファイル解析システムは、以上のように、解析手段が、さらに、上記直線から、所定の距離の範囲内にある第1有意領域を設定する。これにより、発現プロファイルデータの実験誤差を考慮して、表現型に關与するデータの分類を行うことができるという効果を奏する。

40

【0042】

本発明の発現プロファイル解析システムは、以上のように、解析手段が、さらに、上記第3補足データの対応分析結果から、所定の距離の範囲内にある第2有意領域を設定する。これにより、発現プロファイルデータの実験誤差を考慮して、表現型に關係なく常に一定量存在する遺伝子やタンパク質のデータの分類を行うことができるという効果を奏する。

【0043】

本発明の発現プロファイル解析システムは、以上のように、解析手段が、さらに、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を

50

制御する。これにより、膨大な発現プロファイルデータであっても、そのデータに応じて、次元数を制御できる。すなわち、膨大な場合でも、必要な情報量の損失を抑えながら、次元の削減を行うことができ、迅速な解析が可能となる。

【0044】

本発明にかかる発現プロファイル解析システムは、以上のように、上記解析手段は、さらに、対応分析の解析結果を回転可能な画像データに変換することを特徴としている。これにより、上記解析手段が、例えば、解析結果を3次元表示した場合に、その結果を回転可能な画像とする。それゆえ、その解析結果を、モニターなどの表示手段に表示することによって、解析結果を自由に回転移動でき、視覚的に認識しやすいものとして行うことができる

10

本発明の発現プロファイル解析システムは、以上のように、上記発現プロファイルデータは、マイクロアレイ、マクロアレイ、ディファレンシャルディスプレイの少なくともいずれかによって得られたものである。これにより、膨大な量の発現プロファイルデータを、一挙に解析できるハイスループット解析システムを構築することができる。

【0045】

本発明にかかる発現プロファイル解析方法は、以上のように、変換手段が、上記発現プロファイルデータを、 $\arctan(1/\text{ratio})$ を用いて対数変換する変換ステップと、解析手段が、上記変換ステップによって得られた変換値を対応分析する解析ステップとを含んでいる。

【0046】

それゆえ、膨大な数の発現プロファイルデータを、迅速に解析できると共に、発現比が小さい遺伝子やタンパク質も解析対象としているので、標的とする表現型に關与する遺伝子やタンパク質を網羅的に解析できるという効果を奏する。また、発現プロファイルデータを、標的とする表現型に關与するデータと、關与しないデータとに分類できるという効果を奏する。

20

【0047】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記対数変換値に加えて、上記任意の表現型でのみ発現し、その対照となる表現型では発現しないデータである第1補足データと、当該第1補足データと反対の発現様式のデータである第2補足データと、いずれの表現型においても等発現するデータである第3補足データとを用いて対応分析を行うことを特徴としている。それゆえ、表現型に關与するデータの分類の信頼性が向上するという効果を奏する。

30

【0048】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記第1補足データの対応分析結果と、上記第2補足データの対応分析結果とを通る直線を算出することを特徴としている。それゆえ、この直線上のデータは標的とする表現型に關与するデータとして検出し、この直線から外れたデータは關与しないデータとして分類できるという効果を奏する。

【0049】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記直線から、所定の距離の範囲内にある第1有意領域を設定することを特徴としている。それゆえ、発現プロファイルデータの実験誤差を考慮して、表現型に關与するデータの分類を行うことができるという効果を奏する。

40

【0050】

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記第3補足データの対応分析結果から、所定の距離の範囲内にある第2有意領域を設定することを特徴としている。それゆえ、発現プロファイルデータの実験誤差を考慮して、表現型に關係なく常に一定量存在する遺伝子やタンパク質のデータの分類を行うことができるという効果を奏する。

【0051】

50

本発明にかかる発現プロファイル解析方法は、上記の構成に加えて、上記解析ステップは、上記発現プロファイルデータにおける累積寄与率の算出結果に応じて、上記対応分析の次数を制御することを特徴としている。それゆえ、膨大な場合でも、必要な情報量の損失を抑えながら、次元の削減を行うことができ、迅速な解析が可能となる。

【0052】

本発明にかかる発現プロファイル解析プログラムは、以上のように、上記いずれかの発現プロファイル解析システムを動作させるための発現プロファイル解析プログラムであって、コンピュータを上記変換手段および/または解析手段として機能させるためのものである。また、本発明にかかる記録媒体は、以上のように、上記発現プログラム解析プログラムが記録されたコンピュータ読取り可能にしたものである。これにより、プログラムにより本発明にかかる発現プロファイル解析システムをコンピュータで実行させることになるため、コンピュータそのものを本発明にかかる発現プロファイル解析システムとすることができる。その結果、本発明の汎用性を高めることができるとともに、本発明を、通信ネットワーク上で利用することも容易となる。

【発明を実施するための最良の形態】

【0053】

〔実施の形態1〕

本発明の実施の一形態について図1ないし図9に基づいて説明すれば以下の通りである。なお、本発明はこれに限定されるものではなく、特許請求の範囲に示した範囲で種々の変更が可能であり、それぞれの技術的手段を適宜組み合わせ得られる実施の形態についても、本発明の技術的範囲に含まれることはいうまでもない。

【0054】

本発明にかかる発現プロファイル解析システムは、遺伝子および/またはタンパク質の発現プロファイルデータから得られる発現比を \arctan 変換し、その変換値の対応分析に基づいて、標的とする表現型に關与する遺伝子および/またはタンパク質を推定・同定・予測するものである。

【0055】

上記「発現プロファイルデータ」とは、個々の試料、例えば組織、細胞等において発現されている複数の遺伝子および/またはタンパク質の発現パターンを指し、言い換えれば遺伝子および/またはタンパク質の種類とそのそれぞれの発現量（若しくは発現比率）から構成されるデータの集合体を意味する。また、以下では、個々の発現プロファイルデータを、単に、発現データ、遺伝子発現データ、または、タンパク質発現データという。

【0056】

また、上記「表現型」とは、試料（遺伝子および/またはタンパク質）の性格付けに關連する任意の性質を指し、定性的な指標、定量的な指標のいずれもが包含される。例えば、疾病に關連するものでは疾病の名称、原因、進行状況、予後、余命や発症、再発、転移の可能性等が挙げられるが、特に限定されるものではない。

【0057】

また、本発明にかかる発現プロファイルシステムは、マイクロアレイなどによって得られた膨大な量の遺伝子および/またはタンパク質の発現プロファイルデータを効率よく、迅速に処理することが可能であるシステムであり、より具体的には、コンピュータを用いて、発現プロファイル実験、特に網羅的発現プロファイルデータを用いて得られる発現比の対数変換値の対応分析によって、任意の表現型に關与する遺伝子やタンパク質を解析し、その表現型に關与する遺伝子を推定するために好適に利用可能なシステムである。とりわけ、本発明の発現プロファイル解析システムは、発現比が小さい量的形質に關与する遺伝子および/またはタンパク質の検出・推定に好適に利用可能である。

【0058】

「量的形質」とは、解析対象となる表現型（例えば、形質や疾患など）に、多くの遺伝子と環境因子とが複雑に關与し、さらに、それらの遺伝子のわずかな発現量が累積することによって、表現型が連続的に変化することをいう。なお、従来の遺伝子発現プロファイ

10

20

30

40

50

ル解析では、このような遺伝子は、解析対象から排除されていたため、量的形質に關与する遺伝子の発現は困難か、不可能であった。

【0059】

発現プロファイルデータは、マイクロアレイなどによって膨大な数のデータとして一挙に得られるが、このデータの多くは、表現型に關与しない遺伝子またはタンパク質のデータである。そのため、従来では、表現型に關与しているものをある程度絞り込んで、データ解析が行われていた。その結果、発現比の小さい遺伝子は、解析対象とならなかった。

【0060】

具体的には、発現プロファイルデータの解析結果は、大きく以下の(a)~(d)に分類できる。すなわち、

- (a) 標的の表現型または対照となる表現型でのみ特異的に発現するもの。
- (b) 標的の表現型で発現が誘導または抑制され(発現比が相対的に増加または減少)、その表現型の発現に關与するもの。
- (c) 発現量の変化がランダムであり、表現型の発現に關与しないもの。
- (d) 表現型に關係なく等発現するもの。

【0061】

上記(a)~(d)のうち、特に、(a)(b)には、表現型の発現に關与する遺伝子が含まれている。従来の解析手法では、主に(a)および(b)のうち、発現比の変化が大きいもののみを検出しており、(b)のうち発現比の変化が微量なものは、検出対象外となっていた。

【0062】

このため、特に(b)に含まれる、発現比の変化が小さく、かつ、表現型の発現に關与する遺伝子やタンパク質の検出は重要である。

【0063】

本発明は、(a)(b)のうち、特に、(b)に含まれる、発現比の変化が小さく、かつ、表現型に關与するデータを検出するのに好適である。

【0064】

なお、本発明には、発現プロファイル解析システムをコンピュータで実施する場合の発現プロファイル解析方法、および、この解析システムをコンピュータに実行させるプログラム(すなわち、上記発現プロファイル解析システムを動作させるプログラム)、並びに、このコンピュータプログラムを読取可能に記録した記録媒体も含まれる。

【0065】

本実施形態では、遺伝子発現プロファイル解析システムおよび遺伝子発現プロファイル解析方法について説明する。

【0066】

(1) 遺伝子発現プロファイル解析システム

本発明にかかる遺伝子発現プロファイル解析システムは、遺伝子発現プロファイルデータの対数変換値によって遺伝子解析を行うシステムであれば特に限定されるものではない。例えば、図1に示すように、マイクロアレイ51からの網羅的発現プロファイル実験の結果(遺伝子またはタンパク質の発現量)から、遺伝子および/またはタンパク質解析を行う解析システム10aが挙げられる。

【0067】

マイクロアレイ51は、微量の遺伝子(DNA、cDNA、RNAなどのプローブ)が平板上に固定されたものである。本実施形態では、遺伝子発現プロファイル解析を行うので、遺伝子が固定されているが、タンパク質の発現プロファイル解析の場合は、解析対象となるタンパク質と特異的に結合するタンパク質(例えば受容体、酵素など)などの生体物質が固定される。

【0068】

マイクロアレイ51を用いれば、数千以上のDNAやタンパク質に対する反応を同時に実施し、かつ結果の検出も同時に行うことができる。それゆえ、多数の発現プロファイル

10

20

30

40

50

を観察することが可能になる。なお、マイクロアレイ 5 1 の発現プロファイルデータは、任意の解析ソフトウェアによって行えばよい。

【 0 0 6 9 】

なお、マイクロアレイ 5 1 を用いた発現プロファイル実験は、通常、実験誤差や解析結果の信頼性を向上するために、複数回行う。また、マイクロアレイ 5 1 は、マイクロアレイ、遺伝子チップ、プロテインチップ、ディファレンシャルディスプレイ、など、生体物質が基板などに固定されているものであれば特に限定されるものではない。

【 0 0 7 0 】

なお、マイクロアレイ 5 1 による発現プロファイル実験を行っても、表現型に関連する有用な遺伝子は、アレイ上には極めて少なく、大部分は、表現型に関連のない遺伝子である。表現型に関連のない遺伝子の発現は、表現型を直接的に決定しないので、全サンプルを通じて、ランダムで独立した発現レベル、または、類似の発現レベルを示す。

【 0 0 7 1 】

図 1 は、解析システム 1 0 a の概略構成を示すブロック図である。解析システム 1 0 a は、画像読取部 1 1、入力部 1 2、表示部 1 3、画像形成部 1 4、記憶部 1 5、制御部 2 1、変換部 2 2、解析部 2 3、および補足部 3 2 を備えている。

【 0 0 7 2 】

上記画像読取部 1 1 は、マイクロアレイ 5 1 から、プローブにハイブリダイズしたターゲットの蛍光を、信号強度という画像データとして読み取ることで、遺伝子の発現量を検出する。つまり、上記画像読取部 1 1 は、解析用変量としてマイクロアレイ 5 1 から得られる発現プロファイルデータを、遺伝子の発現量に比例して変化する信号強度として検出して遺伝子発現プロファイル解析システムに入力する入力手段である。

【 0 0 7 3 】

上記画像読取部 1 1 としては、具体的には、例えば、蛍光スキャナー等が好適に用いられるが、特にこれに限定されるものではなく、ターゲットを標識している色素の種類に応じて、適切な構成の画像読取部 1 1 を選択すればよい。

【 0 0 7 4 】

上記入力部 1 2 は、上記解析システム 1 0 a の動作に関わる情報を入力可能とする。具体的には、キーボードやタブレット等、従来公知の入力手段を好適に用いることができる。また、マイクロアレイ 5 1 からの得られる遺伝子の発現量は、必ずしも上記画像読取部 1 1 から読み取られるものではなく、例えば、別の読取手段等で読み取られた後に具体的な数値データに変換されたとすれば、上記入力部 1 2 から上記解析システム 1 0 a に入力することもできる。また、入力部 1 2 に、公知の遺伝子発現プロファイルデータを入力することによって、そのデータの解析を行うことも可能である。

【 0 0 7 5 】

つまり、本実施形態では、試料となる遺伝子群から、網羅的発現プロファイル実験により発現量のデータが得られればよく、解析システム 1 0 a への入力の動作としては、画像読取部 1 1 による信号強度の直接読み取りに限定されるものではない。それゆえ、本発明においては、入力手段として、上記画像読取部 1 1 および入力部 1 2 の少なくとも一方を備えていることが好ましいが、入力手段としては、上記画像読取部 1 1 や入力部 1 2 に限定されるものではなく、その他の入力手段を備えていても良い。

【 0 0 7 6 】

上記表示部 1 3 は、マイクロアレイ 5 1 からの信号強度の読み取りや、読み取った信号強度の解析等を含む、上記解析システム 1 0 a の動作に関わる情報や解析結果等の各種情報を表示する。具体的には、公知の CRT ディスプレイや、液晶ディスプレイ等といった各種表示装置を好適に用いることができるが特に限定されるものではない。表示部 1 3 では、例えば、解析結果を 3 次元表示し、そのグラフを回転可能なようにすることによって、認識しにくい解析結果を、視覚的に認識しやすくすることが可能である。

【 0 0 7 7 】

上記画像形成部 1 4 は、上記表示部 1 3 で表示可能な各種情報を P P C 用紙等の記録材

10

20

30

40

50

に記録（印刷・画像形成）する。具体的には、公知のインクジェットプリンタやレーザープリンタ等の画像形成装置が好適に用いられるが特に限定されるものではない。

【0078】

なお、上記表示部13と画像形成部14とは、まとめて出力手段と表現することもできる。すなわち、表示部13は、各種情報をソフトコピーで出力する手段であり、画像形成部14は、各種情報をハードコピーで出力する手段である。したがって、本発明で用いられる出力手段としては、上記表示部12や画像形成部13に限定されるものではなく、その他の出力手段を備えていても良い。

【0079】

上記記憶部15は、上記解析システム10aで利用される各種情報（制御情報、解析結果、その他情報等）を記憶する。具体的には、例えば、RAMやROM等の半導体メモリ、フレキシブルディスクやハードディスク等の磁気ディスクやCD-ROM/MO/MD/DVD等の光ディスクのディスク系、ICカード（メモリカードを含む）/光カード等のカード系等、従来公知の各種記憶手段を好適に用いることができる。

【0080】

上記制御部21は、解析システム10aの動作を制御する。具体的には、図1の点線の矢印で示すように、画像読取部11、入力部12、表示部13、画像形成部14、記憶部15、変換部22、解析部23、および補足部32の各手段に対して、上記制御部21から制御情報が出力される。この制御情報に基づいて上記各手段が連携して動作することで、上記解析システム10a全体が動作する。また、制御部21に対しては、入力部12から解析システム10aを動作させるための指示情報も入力可能となっているので、図1では、制御情報のやりとりを示す点線の矢印は双方向となっている。

【0081】

変換部22は、画像読取部11で読み取られた発現プロファイルデータ、または、入力部12に入力された発現プロファイルデータを、対数変換する。

【0082】

本発明では、変換部22における対数変換指標が、『発現比の逆数の逆正接関数（ $\arctan(1/\text{ratio})$ ）』を用いることを特徴としている。ここで、ratioは、任意の表現型での遺伝子発現量と、比較対照となる表現型での遺伝子発現量の比なお、従来の対数変換指標としては、『発現比の対数（ $\log(\text{ratio})$ ）』が用いられていた。また、従来の対数変換指標では、発現比が1（等発現）付近の遺伝子については、表現型に関与しない遺伝子とみなされていたため、例えば、量的形質などのように、この付近に存在する遺伝子については、検出されなかった。また、対数の底がバラバラであったため、得られた値を直接比較できなかった。

【0083】

ここで、上記 $\arctan(1/\text{ratio})$ において、「 $1/\text{ratio}$ 」は、 $\log_e(\text{ratio})$ の微分係数であり、 $- \sim +$ の値をとる。本発明の遺伝子発現プロファイル解析システムでは、対応分析を用いるが、この対応分析は、負の値を取り扱うことはできない。そこで、「 $1/\text{ratio}$ 」を \arctan 変換することによって、全ての値を正の値にすることが可能となる。

【0084】

図3は、本願による対数変換と、従来の対数変換とを比較したグラフである。すなわち、図3は、本願の対数変換指標（ \arctan 変換値）と従来の対数変換指標（種々の底の \log 変換値）とを、遺伝子発現比に対してプロットしたグラフである。同図に示すように、 \arctan 変換値の場合、従来の対数指標とは、全く異なるグラフとなる。特に、 \arctan 変換値の場合、発現比が比較的小さい場合、例えば、0.1倍～10倍の付近の変動であっても、高精度に検出できる。

【0085】

すなわち、 \arctan 変換値を用いた場合のグラフは、以下の特徴を有している。第1に、発現比が1倍（等発現）のときの 45° を中心として、発現比が、上昇するにつれ 0° に、減少するにつれ 90° に変化する。第2に、発現比が1倍を中心として約1/10倍から10

10

20

30

40

50

倍の範囲のとき、arctan変換値の変化率は、従来指標の変化率と比較して著しく大きい。すなわち、1倍付近のわずかな遺伝子の発現レベルの変動に対しても検出力が高いことを意味する。

【0086】

したがって、有意性があるにもかかわらず、従来は見落とされていた等発現付近の遺伝子であっても、本発明では確実に検出できる。それゆえ、本発明の遺伝子発現プロファイル解析システムによって得られたデータは、信頼性の高いデータである。したがって、本発明によれば、従来は見落とされていた、有意性のある遺伝子を検出できる可能性が極めて高い。

【0087】

また、従来指標(log変換値)では- から+ のいかなる範囲を取り得るが、arctan変換値では0度から90度の有限の範囲をとる。発現比が、著しく高い、もしくは、低い遺伝子は、統計解析の結果、多くの場合、有意性が認められるため、これらの発現比に対する指標の値は、必ずしも- や+ の近傍といった極値をとる必要は無い。また、そのような外れ値の存在は、統計解析の結果、それよりも弱い作用を持つ遺伝子の検出を困難とする場合がある。そこで、有限の値をとるarctan変換値を用いることで、極端に高い、もしくは、低い値の出現頻度は少なくなり、解析における外れ値の影響を抑えることができるという利点も持つ。

【0088】

解析部23は、変換部22における遺伝子発現プロファイルの変換値および補足部32のデータに基づいて、対応分析を行う。解析部23の対応分析は、例えば、大型計算機に限らず、一般的なパーソナルコンピュータ、Windows(登録商標)マシンやUnix(登録商標)、Linux、マッキントッシュ(MacOS)などによって行うことができる。対応分析を行うと、互いに関連のある項目が、近くに集合する。これにより、サンプルの表現型に關与する遺伝子は、近くに集合し、関連しない遺伝子と区別することができる。

【0089】

なお、後述のように、対応分析は、フリーの統計パッケージRとそのR上で対応分析や主成分分析を行うためのライブラリmultivを用いて行うことができる。なお、『「記述的多変量解析法」、日科技連、大隈ら』にも、FORTRANのプログラム、SASなどによっても、対応分析は可能である。また、対応分析は、数量化III類とも呼ばれ、この名前でもパッケージされている。

【0090】

解析部23は、補足部32のデータを使用することなく、変換部22の変換値のみに基づいて、解析することも可能である。ただし、より正確な検出を行うためには、変換部22と補足部32のデータを使用することが好ましい。

【0091】

また、解析部23は、遺伝子発現プロファイル解析速度や精度の向上のために、UDL(直線)、UDR(第1有意領域)、および累積寄与率に応じて対応分析スコアの次元などを設定する。UDLおよびUDRについては後述するが、UDLを設定することにより、UDL上には標的とする表現型に關与する遺伝子が位置する。UDRは、実験による誤差を考慮してUDLから一定の距離をおいて算出した有意領域であり、UDR内には、解析対象となる遺伝子が位置する。

【0092】

ここで、解析部23で行われる対応分析について、簡単に説明する。対応分析は、対応分析とは、主成分分析と同様に、n次元のデータを説明するための主軸を決定する解析手法である。具体的には、本実施形態では、遺伝子発現プロファイルデータから、表現型(形質など)の違いを説明できる1つ、もしくは、複数の主軸を求める。ここで、1つの主軸だけで、表現型の違いが説明されるのであれば、それは1次元であり、その主軸の寄与率は100%となる。また、1つの主軸では表現型の変化を説明できず、例えば、表現型の違いの説明に、2つの主軸(第1・第2主軸)が必要な場合には、第1・第2主軸によ

10

20

30

40

50

って説明がなされる割合（寄与率）が、それぞれ、例えば、70%と30%などとなる。

【0093】

つまり、上記「寄与率」とは、各主軸によって、表現型の変化を説明がなされる割合を示し、上記「累積寄与率」とは、その寄与率の和を示している。このとき、第1主軸の寄与率は、第2主軸の寄与率と等しい、もしくは、それ以上となる。同様に、第3、第4主軸となるにしたがって、寄与率は低下する。第1および第2主軸によって表現型の違いの説明が可能な場合、解析結果を示す図は、1次元もしくは2次元プロットで描くことができる。また、表現型の違いの説明に、第3主軸までを必要とする場合には、解析結果を示す図は、3次元図までのプロットで描くことができる。このように、対応分析では、累積寄与率が100%となるまで、次元の数（すなわち、主軸の数）が増えていく。

10

【0094】

なお、上記寄与率は、各主軸に与えられる固有値から算出する。具体的には、全主軸の固有値の和に対する各主軸の固有値の比が、その主軸の寄与率となる。例えば、対応分析によって、表現型の変化を説明するために、10次元までの主軸（第1～第10主軸）が得られたとき、各主軸に対して固有値が与えられる。そして、この各主軸に対する固有値の総和に対する各主軸の固有値の割合が寄与率となり、第1主軸から第10主軸まで順次寄与率の和を求めていったものが累積寄与率となる。

【0095】

ここで、極端な例として、対応分析の計算では、10次元まで算出され、5次元までの累積寄与率が98%であったとする。この場合、残りの6次元以降は、2%しか表現型の違いを説明していないことになる。よって、6次元以降以降の解析結果は、無視することも可能である。

20

【0096】

すなわち、対応分析の結果、3次元まで算出された場合（主軸が3つ）、プロットは1次元、2次元、3次元図で表すことができる。また、累積寄与率は、3次元の場合で全体の100%となり、3次元のプロット図で、表現型の違いを完全に説明することができる。しかし、対応分析の結果、4次元以上の主軸が算出された場合、4次元以上のプロットは、実際には不可能である（数学的には可能であるが、通常のプロットでは行わない）。このため、全次元をもつての視覚化はできない。しかし、例えば、4次元以上の主軸が算出され、第3主軸までの累積寄与率が90%で、それ以降の主軸の寄与率が10%といった場合には、3次元図でも、全体の90%は説明できる。すなわち、90%の精度での判定が可能となる。この場合、残りの10%に含まれる遺伝子は、図では説明できないので、4次元図を3次元に落とした際に、直線UDLやUDRから外れるプロットも、いくつか出現する可能性がある。

30

【0097】

このように、遺伝子プロファイル解析では、対応分析によって、試料の数と同じ次元が得られるが、累積寄与率を算出することによって、その次元を低くすることができる。これにより、データ処理速度は速くなり、迅速な解析が可能となる。

【0098】

補足部32は、1つの表現型でのみ特異的に発現する遺伝子群のデータと、全ての表現型で等発現する遺伝子群のデータとを、解析部22に人為的に付加する。具体的には、例えば、2つの表現型（ある表現型と、その対照となる表現型）を比較する場合、(i)一方の表現型では著しく高い発現比を有する一方、他方の表現型では発現しない遺伝子（発現比0）（第1補足データ）、および、(ii)その逆の発現比を有する遺伝子（第2補足データ）、(iii)いずれの表現型でも発現比1を有する遺伝子のデータを付与する（第3補足データ）。このうち、(i)・(ii)は、一方の表現型でのみ特異的に発現する遺伝子であり、表現型を支配する遺伝子を検出するのに用いる。(iii)は、ハウスキーピング遺伝子の検出に用いる。このような(i)～(iii)のデータを付加することによって、これらのデータに基づいて、実際のマイクロアレイデータの変換値の解析を行うので、より正確な解析結果が得られる。なお、補足部32で補足されるデータは、変換部22と

40

50

同様、 $\arctan(1/\text{ratio})$ 変換したものである。つまり、補足部 3 2 のデータは、変換部 2 2 に入力されてもよい。

【 0 0 9 9 】

解析部 2 3 は、補足部 3 2 から付与される上記第 1 補足データおよび第 2 補足データの対応分析結果を通る直線を算出してもよい。上記第 1 および第 2 補足データは、それぞれの表現型に特異的に発現する遺伝子の発現データであるので、上記の直線上には、表現型に關与する遺伝子がプロットされる。すなわち、この直線は、前述の U D L となる。それゆえ、第 1 および第 2 補足データから算出した直線を算出しておけば、その直線上にプロットされた遺伝子は、表現型に關与すると推定できる。なお、U D L は、補足部 3 2 からの第 1 および第 2 補足データの対応分析結果に基づいて作成してもよいし、発現比（あるいは発現量）が公知のデータの場合、変換部 2 2 から得られた変換値に基づき、表現型に關与する遺伝子のプロットを結んで作成してもよい。いずれにしても、U D L 上には、標的とする表現型に關与する遺伝子がプロットされる。

10

【 0 1 0 0 】

また、解析部 2 3 は、解析結果を、回転可能な画像データに変換することが好ましい。これにより、解析結果を表示部 1 3 の画面に表示し、その結果を自由に回転できる。それゆえ、プリントアウトした場合に分かり難い高次元の解析結果でも、画面上で認識しやすくなる。このデータ変換には、例えば、後述の実施例のように、本発明者らによって確立されたクリックブルビューアー-e-GRED(clickable viewer e-GRED) (Copyright (c), 矢野健太郎・清水顕史 All Rights Reserved) を用いるのが好適である。

20

【 0 1 0 1 】

このように、解析システム 1 0 a では、図 1 の実線の矢印で示すように、画像読取部 1 1 から得られた発現プロファイルデータが、変換部 2 2 に出力されて、対数変換値に変換され、次に、その対数変換値と、補足部 3 2 にて付与される補足データとが解析部 2 3 に入力され、対応分析による解析が行われる。そして、最終的に、解析部 2 3 の解析結果が表示部 1 3 および/または画像解析部 1 4 に出力されることになる。

【 0 1 0 2 】

すなわち、解析システム 1 0 a は、図 2 のフローチャートにしたがって、発現プロファイルデータの解析を行う。なお、解析方法の詳細については、後述の (2) 遺伝子発現プロファイル解析方法で具体例を挙げて説明する。

30

【 0 1 0 3 】

まず、前段階として、網羅的発現プロファイル実験を実施する。具体的には、前述したように、特定の生物のゲノムに含まれる全ての遺伝子またはその一部をプローブとして用いたマイクロアレイ 5 1 に対して、蛍光色素で標的したターゲット DNA (以下、ターゲットと略す) をハイブリダイズさせる。上記マイクロアレイ 5 1 は、従来公知の手法で作製することができ、その作製手法については特に限定されるものではない。

【 0 1 0 4 】

上記前段階としての網羅的発現プロファイル実験は、1 回のみ実施されてもよいが、通常は複数回実施される。そこで、ステップ 1 1 (以下、ステップを適宜 S と略す) として、ターゲットの蛍光を、画像読取部 1 1 で信号強度として測定 (検出) し、解析用変量となる遺伝子の発現量のデータを入力する (発現プロファイルデータ入力ステップ)。

40

【 0 1 0 5 】

上記 S 1 1 は、実施された全ての実験の結果からデータを入力し終わるまで繰り返される。それゆえ、S 1 2 として、全ての発現量のデータが入力されたか否かを判定し、入力されていれば、S 1 3 に進む一方、入力されていなければ、S 1 1 に戻る。

【 0 1 0 6 】

次に、S 1 3 として、変換部 2 2 により、発現プロファイルデータ (発現量) の変換を実施する (変換ステップ)。具体的には、変換部 2 2 では、前述したように、発現比の対数変換指標として、(\arctan 変換値) を算出する。この対数変化指標により、発現比が小さい遺伝子も検出可能となる。

50

【0107】

次に、S14として、補足部32により、後続する解析速度および解析精度の向上のために、補足データ（前述の第1～第3補足データ）が付与される（補足ステップ）。この補足データは、表現型の発現に關与する遺伝子の検出およびハウスキーピング遺伝子の検出に用いるものである。

【0108】

続いて、S15として、変換部22で得られた対数変換値（arctan変換値）および補足部32で得られた補足データを用いて、解析部23にて対応分析による解析を行う（解析ステップ）。この解析ステップでは、累積寄与率による対応分析の次元設定や、UDL・UDRの設定を適宜行ってもよい。

10

【0109】

その後、S16として、解析結果を出力する。具体的には、表示部13に表示したり、画像形成部14でプリントアウト（印刷）したりする（解析結果出力ステップ）。

【0110】

以上のように、解析システム10aは、変換部21および解析部23が設けられていることを特徴としている。これにより、発現プロファイルデータの新規対数変換指標として、arctan変換値を用いるので、発現比が1倍付近のわずかな発現比の変化も高感度に検出することが可能である。したがって、発現プロファイルデータの全てを解析対象とし、発現比の大小にかかわらず、表現型に關与する遺伝子を検出することができる。この解析では、膨大なデータを処理するにもかかわらず、特別な計算機が不要であり、通常のコンピュータ等での解析が可能である。さらに、ハウスキーピング遺伝子の検出も可能である。それゆえ、膨大なデータの迅速な解析が可能であるとともに、解析結果の精度も向上する。

20

【0111】

なお、以上説明した本実施の形態における解析システム10aは、以上説明したS11～S17までのステップを含む網羅的発現プロファイル解析方法を機能させるためのプログラムにより、コンピュータで実現されるようになっていてもよい。

【0112】

上記プログラムはコンピュータで読み取り可能な記録媒体に格納されていればよい。具体的には、図1に示す記憶部15、具体的には、例えばROMのようなものそのものがプログラムメディアであってもよいし、記憶部15として、プログラム読み取り装置が設けられている場合には、そこに記録媒体を挿入することで読み取り可能なプログラムメディアであってもよい。上記プログラムメディアとしては、記憶部15の具体例として挙げた公知の構成を好適に用いることができる。

30

【0113】

何れの場合においても、格納されているプログラムは制御部21がアクセスして実行させる構成であってもよいし、プログラムを読み出し、読み出されたプログラムを、図示しないプログラム記憶エリアにダウンロードして、そのプログラムを実行する方式であってもよい。このダウンロード用のプログラムは予め記憶部15等に格納されているものとする。また、上記記録媒体に格納されている内容はプログラムに限定されるものではなく、例えばデータであってもよい。

40

【0114】

なお、図1では、実際にマイクロアレイ51による発現プロファイル実験を行って発現プロファイルデータを得ているが、発現プロファイルデータは、公知となっている発現プロファイルデータを使用してもよい。この場合、入力部12に、そのデータを入力し、そのデータを、変換部22にてarctan変換値に変換すればよい。

【0115】

また、解析システム10aは、遺伝子発現プロファイル解析システムであるが、マイクロアレイの代わりに、例えば、プロテインアレイを適用し、タンパク質の発現比のarctan変換値を用いて、対応分析することにより、タンパク質の発現プロファイル解析システム

50

とすることも可能である。

【0116】

また、解析システム10aは、DNAマイクロアレイと、プロテインアレイ、プロテオーム解析装置や、タンパク質相互作用解析装置などのタンパク質を解析する装置とを組み合わせた構成であってもよい。遺伝子によっては、細胞内の発現量と、タンパク質の生産量とが一致しない場合がある。また、タンパク質同士の相互作用や、翻訳後修飾などは、遺伝子の発現量を解析するだけでは、把握できない場合がある。したがって、タンパク質解析装置を備えることによって、遺伝子の発現量に加えて、タンパク質の発現量も反映させて、表現型に関連する遺伝子およびタンパク質の解析を行うことが可能である。

【0117】

それゆえ、例えば、疾患に関連する遺伝子およびタンパク質を推定できるので、新規医薬品の標的を探索することが可能となる。

【0118】

プロテオーム解析装置は、特に限定されるものではないが、例えば、2次元電気泳動を用いた解析や、質量分析、表面プラズモン解析を行えるような装置が挙げられる。

(2) 遺伝子発現プロファイル解析方法

本発明の遺伝子発現プロファイル解析方法は、遺伝子発現プロファイル解析システムを用いることによって好適に実施することができる。

【0119】

すなわち、本発明の遺伝子発現プロファイル解析方法は、遺伝子発現プロファイルデータの対数変換値によって、遺伝子解析を行う遺伝子発現プロファイル解析システムの遺伝子発現プロファイル解析方法であって、変換部22が発現プロファイルデータを、 $\arctan(1/\text{ratio})$ を用いて対数変換する変換ステップと、解析部23が変換ステップで得られた変換値を対応分析する解析ステップとを含んでいる。

【0120】

ここで、本実施形態の遺伝子発現プロファイル解析方法について、図4に示した遺伝子発現プロファイルのデータの対応分析を例に挙げて説明する。図5(a)には図4のデータの対応分析の結果を、図5(b)には、比較のために、同データの主成分分析の結果を示した。

【0121】

図4は、2つの表現型AおよびB(Sample)を有する表現型を調査対象とし、24個の遺伝子を各表現型について3サンプルずつ(計6サンプル)の発現量を測定したデータである。なお、表1は、図4のデータを対応分析した結果の一例を示している。

【0122】

【表1】

次元	固有値	寄与率	累積寄与率
1	0.0052	0.33	0.33
2	0.0025	0.16	0.49
3	0.0021	0.13	0.63
4	0.0017	0.11	0.73
5	0.0016	0.10	0.83
6	0.0014	0.09	0.92
7	0.0012	0.08	1.00
計	0.0157	1.0000	

【0123】

図4において、遺伝子D1~D6は表現型Aに対して表現型B側での発現量が多い遺伝子群(表現型B側で誘導された遺伝子群)であり、遺伝子U1~U6は表現型Bに対して

10

20

30

40

50

表現型 A 側での発現量が多い遺伝子群（表現型 A 側で誘導された遺伝子群）である。すなわち、これらの遺伝子 D 1 ~ D 6 および U 1 ~ U 6 の 1 2 個の遺伝子は、表現型が A であるか B であるかの決定に關与する遺伝子であり、本発明において標的とする表現型に關与する遺伝子である。

【 0 1 2 4 】

一方、遺伝子 Unrelated 1 ~ Unrelated 6 は、それぞれ、6 サンプルでランダムな発現比を示す遺伝子群である。すなわち、遺伝子 Unrelated 1 ~ Unrelated 6 は、表現型の決定に關与しない遺伝子群である。

【 0 1 2 5 】

また、遺伝子 HK 1 ~ HK 3 は、ハウスキーピング遺伝子であり、各遺伝子ともに、全 6 サンプルが同一の発現比を示している。

【 0 1 2 6 】

遺伝子 Supplement 1 ~ Supplement 3 は、表現型 A または B の遺伝子発見を容易にするために人為的に付加されるデータセットである。具体的には、遺伝子 Supplement 1（第 1 補足データ）は、表現型 A において著しく高い発現比をもつ一方で、表現型 B では発現していない遺伝子（発現比 0）である。Supplement 2（第 2 補足データ）は、Supplement 1 と逆の発現パターンの遺伝子であり、表現型 B において著しく高い発現比をもつ一方で、表現型 A では発現していない遺伝子（発現比 0）である。すなわち、これら 2 つの遺伝子は、いずれか一方の表現型でのみ特異的に発現している遺伝子であり、表現型を支配する遺伝子を検出する際の糸口となる。

【 0 1 2 7 】

また、Supplement 3（第 3 補足データ）は、全 6 サンプルともに発現比 1 をもつ遺伝子であり、ハウスキーピング遺伝子の同定の糸口となる遺伝子である。

【 0 1 2 8 】

この遺伝子 Supplement 1 ~ Supplement 3 のデータセットは、前述の補足部 3 2 に格納されており、発現プロファイルデータの分類を容易にする。

【 0 1 2 9 】

表 1 に示すように、各遺伝子は、7 次元の座標（スコア）を有している。したがって、7 次元プロットを図示できるのであれば、その図は、各遺伝子による表現型の変化を 1 0 0 % の精度で示すことが可能である。すなわち、高精度の判定結果を得るためには、各遺伝子のどの遺伝子が、UDR 内に含まれるかの判定は、この累積寄与率が 1 0 0 % である 7 次元での距離を求めることによって行う。しかし、図示が可能な第 3 主軸までの累積寄与率が 6 3 % ほどであっても、図 5（a）のように、UDR からはみ出る遺伝子のプロットは少なく、第 3 主軸までの 3 次元プロットでも、判定精度に大きな問題はなく、視覚化において大きな問題は見られなかった。ただし、仮に、第 3 主軸までの累積寄与率が、3 0 % などと極端に低いなら、UDR から完全に飛び出してしまうプロットも出てくるため、良好な視覚化に影響が生じるおそれがある。

【 0 1 3 0 】

したがって、図示が可能な第 3 主軸までの累積寄与率によって判定する場合、例えば、累積寄与率が 6 0 % 程度以上であれば、UDR からはみ出るようなプロットは少なくなり、良好な視覚化には大きな影響を及ぼさないと考えられる。しかしながら、表現型に關与する遺伝子の判定を、図示が可能な 3 次元までで行うのではなく、累積寄与率が 1 0 0 % となる n 次元で行うことによって、UDR からはみ出るプロットが存在することなく、より一層良好な視覚化が可能となる。

【 0 1 3 1 】

図 5 には、以上のような図 4 に示す遺伝子発現プロファイルのデータセットを、解析部 2 3 にて対応分析（図 5（a））した結果を示した。なお、比較のため、同データを従来通り主成分分析（図 5（b））した結果も示した。

【 0 1 3 2 】

主成分分析では 6 次元、対応分析では 5 次元までの主軸（Factor）が出力されたが、解

10

20

30

40

50

析部 2 3 で検出された 3 次元までの累積寄与率は、それぞれ、90.3%と95.0%であったため、図 5 では、3 次元までの累積寄与率のデータを図示している。このように、累積寄与率が比較的高い場合には、低次元での累積寄与率を適用できる。言い換えると、必要な情報量の損失を抑えながら、次元の削減を行うことができる。これにより、データ処理速度を上昇させるとともに、3 次元程度であれば図示した結果も視覚的に認識しやすくなる。

【0133】

まず、図 5 (b) に示した、従来の方法である主成分分析の結果についてみると、遺伝子 D 1 ~ D 6 と、遺伝子 U 1 ~ U 6 とは、主軸 1 (Factor 1) によって、正負に分離されている。しかしながら、ハウスキーピング遺伝子 (H K 1 ~ H K 3) と、遺伝子 Supplement 3 とは、異なる座標にプロットされている。そのため、主成分分析では、ハウスキーピング遺伝子を同定することは困難である。また、表現型に関与しない遺伝子群 (遺伝子 Unrelated 1 ~ Unrelated 6) も、表現型 A または B のいずれかと同じ座標にプロットされるため、同定が困難である。

10

【0134】

これに対して、図 5 (a) に示す対応分析では、表現型に関与する遺伝子を分離できるだけでなく、ハウスキーピング遺伝子の同定、さらには、表現型に関与しない遺伝子の分離も可能となる。

【0135】

すなわち、遺伝子 D 1 ~ D 6 と、遺伝子 U 1 ~ U 6 とは、主軸 1 (Factor 1) によって、正負に分離されているという点では、図 5 (a) 一致している。しかし、図 5 (a) では、遺伝子 D 1 ~ D 6 および遺伝子 U 1 ~ U 6 は、直線上に位置している。さらに、この直線は、Supplement 1 および Supplement 2 を結ぶ直線と一致している (Supplement 1 および Supplement 2 を結ぶ直線が「UDL」である)。

20

【0136】

すなわち、対応分析では、標的とする表現型 A および B に関連する遺伝子は、この UDL 上に位置することになる。したがって、理論上は、直線 UDL に沿ってプロットされた遺伝子は、標的の表現型に関与する遺伝子と推定できる。なお、この UDL は、解析部 2 3 で作成される。

【0137】

しかしながら、実際の生物学的測定値は、期待値からの偏差を伴うため、標的の表現型に関与する遺伝子が、完全に UDL に沿ってプロットされとは限らない。このため、UDL からの有意距離を定義し、設定することが好ましい。これにより、UDL から有意距離の範囲 (第 1 有意領域) 内にあるものを、標的とする表現型に関連する遺伝子と推定することができる。すなわち、生物学的測定による実験誤差を考慮した上で、標的の表現型に関与する遺伝子を推定することが可能となる。それゆえ、より実用性の高い遺伝子発現プロファイル解析法を提供することができる。

30

【0138】

UDL からの有意距離は、補足部 3 2 によって、例えば、カイ二乗距離によって定義することができる。具体的には、有意距離を、UDL から任意の有意水準の距離として定義する。その結果、例えば、第三主軸までを用いた視覚化では、3 次元で図示されるので、図 5 (a) に示すように、UDL から有意距離にある領域 (第 1 有意領域) は、UDL を中心軸とする、統計的に有意なカイ二乗距離を半径とする円柱となる。これにより、この円柱の内部に位置する遺伝子は、標的の表現型に関与する遺伝子と推定し、外部に位置する遺伝子は標的の表現型に関与しない遺伝子と推定できる。すなわち、図 5 (a) に示すように、表現型 A または B に関連する遺伝子群 D 1 ~ D 6 および U 1 ~ U 6 は、UDL の内部に位置し、関与しない遺伝子群 Unrelate 1 ~ 6 は、UDL の外部に位置している。

40

【0139】

なお、カイ二乗距離によって有意領域を定義するための有意水準は、適宜設定すればよく特に限定されるものではない。一般的な統計で用いられる有意水準 95% 程度に設定す

50

ることが好ましい。また、有意領域は、対応分析で出力されたすべての次元に対する距離を算出して表示される。図5(a)は3次元データであるので、有意領域も幾何学的に3次元の円柱として表示されている。

【0140】

また、カイ二乗距離以外にも、例えば、ユークリッド距離などによっても、有意距離を定義することが可能である。また、カイ二乗距離の算出は、必ずしも3次元に限ったものではなく、累積寄与率が十分に得られる次元数で行うことで、より正確な判定が可能である。ただし、より良好な視覚化を実現するには、対応分析で出力されたすべての次元、すなわち、累積寄与率が100%の次元で、カイ二乗距離などを計算することが好ましい。

【0141】

なお、図5(b)では、ハウスキーピング遺伝子(HK1~HK3、Supplement3)も、UDRの内部に位置している。図5(b)では、ハウスキーピング遺伝子(HK1~3およびSupplement3)が、1点に集中してプロットされている(Supplement3は重なって視覚化できていないだけである)。なお、ハウスキーピング遺伝子も、UDR内の直線UDL上にプロットされる。しかし、ハウスキーピング遺伝子は、UDL上の一点に集中してプロットされるため、標的の表現型に關与する遺伝子と区別して推定することが可能である。なお、ハウスキーピング遺伝子の推定にも、前述のように、補足部32にて、遺伝子Supplement3を中心とし、統計的に有意なカイ二乗距離を半径とする球を用いることができる。すなわち、Supplement3から有意距離にある領域が、第2有意領域である。この球の内部に存在する遺伝子は、総計学的に有意なハウスキーピング遺伝子とみなすことができる。これにより、実験誤差を考慮して、ハウスキーピング遺伝子を推定することが可能となる。なお、表現型に關連しない遺伝子およびハウスキーピング遺伝子を推定するために、対応分析から得られたスコアには、マイクロソフトエクセル(登録商標)などの数値コンピュータ言語を用いることができる。

【0142】

なお、前述のカイ二乗距離は、式(1)から算出した。

【0143】

【数1】

$$\sqrt{\chi^2 / \sum_i f_i} \quad \dots \dots (1)$$

【0144】

ここで f_i は、 i 番目(i は自然数)のサンプル遺伝子とカイ二乗(χ^2)の自由度との新規指標($\arctan(1/\text{ratio})$)であり、対応分析によって得られるファクターの数に等しい。図4の解析では、Supplement1遺伝子の6個のサンプルの指標は、ベクトル(0,0,0,90,90,90)、および $n=5$ (n は対応分析の結果、累積寄与率が100%となるまでに得られた次元(主軸)の数であり、カイ二乗の自由度)である。それゆえ、カイ二乗は有意水準95%で11.07であり、数(1)に代入すると、 $df=5$ 、Supplement1遺伝子からの有意距離は、 $(11.07/270)^{1/2}=0.2024$ となる。同様にして、Supplement2遺伝子からの有意距離が得られる。これらの有意距離は、UDRの範囲として使用できる。

【0145】

UDLは、Supplement1とSupplement2の n 次元スコアからの線形補間法(内挿法)によって規定される。UDLとある遺伝子との距離が、上記有意距離を越えている場合(すなわち、UDRの外部)、その遺伝子は、表現型に關与しない遺伝子と推定できる。表現型に關与しない遺伝子のみではなく、ハウスキーピング遺伝子も、Supplement3とその有意距離によって推定できる。

【0146】

なお、UDR内に位置するかの判定は、対応分析で得られた累積寄与率が100%である n 次元での距離を用いて行い、図示のための次元を落とした3次元では行わない。その理由は、次元を3次元まで落とすと累積寄与率が低下する、すなわち、表現型の違いを第

10

20

30

40

50

3主軸まででは100%説明しきれないためである。また、3次元に落として図示した際に、n次元ではUDR内に位置するのに、3次元ではUDRからやや飛び出るものも出てくる場合もあるためである。なお、このUDRからやや飛び出る程度は、第3主軸までの累積寄与率の大小に依存する。

【0147】

図9は、遺伝子発現プロファイル解析を、解析の手段として、「R」を用いた場合の解析プログラムの一例である。図9は、対応分析や主成分分析を行うためのソフトとして、デフォルトで最大7次元まで求めるRの統計ライブラリmultivを示している。このライブラリでの主成分分析と対応分析を行うためのコマンドは、それぞれ、PCAとCAである。図9で、入力する値は、遺伝子発現プロファイル(例えば、アレイで求めた遺伝子発現量、各行に遺伝子、各列にサンプル[個体、被験者など])であり、一般的なアレイの実験装置が出力する実測値データである。

10

【0148】

Data <- as.matrixの行で、その実測値データのファイルをPCメモリに読み込み、lnの行でarctan(1/ratio)の変換を行い、次に、CAOUTの行で対応分析を行う(ここで、DataやINはユーザーが任意に決める変数名であり、この解析例はあくまでも一例に過ぎない)。次に、対応分析や主成分分析を行うためのコマンドCAとPCAで、デフォルトでは最大7次元までの解析が行われる。もし、7次元までで十分な累積寄与率が得られず、8次元以上で計算を行うのであれば、その変更も可能である。

【0149】

n次元までの算出が終わると、各遺伝子と各サンプルに対して、n次元のスコア(座標)が出力される。例えば、6次元で累積寄与率が100%であれば、各遺伝子と各サンプルは、6つの軸に対するスコア(座標値)を有することになる。

20

【0150】

「R」のコマンドのwrite.tableでのrproj、rproc、evalsは、それぞれ、行(遺伝子)、列(サンプル)のn次元スコア(座標値)と各主軸の固有値とを、PCメモリからテキスト・ファイルに書き出すためのものである。これらの遺伝子とサンプルとのスコア(座標値)を使ってn次元距離を求め、UDRに含まれる遺伝子の判定やプロットなどを行う。固有値は、前述のように、各主軸の寄与率と累積寄与率を求めるために用いる。

【0151】

7つの主軸の対応分析および主成分分析(PCA)は、統計的コンピュータ言語のRおよびその包括的な多変量である(図5・9)。図9の中央下部の、記号(>)のあるコマンドは、Rコマンドの使用例を示している。入力ファイル“example.txt”は、図4(この場合、行列は、24行、6列である)から、タブで区切った発現比のみを有している必要があるが、コマンドを変更すれば、カンマでも空白区切りでもよい。出力される6ファイルは、対応分析および主成分分析における、行列および固有値のスコアを有している。

30

【0152】

以上のように、本発明にかかる遺伝子発現プロファイル解析システムおよび解析方法によれば、遺伝子発現比の新規指標としてarctan(1/ratio)を用いることによって、同等の量的形質における表現型に影響を及ぼす可能性のある遺伝子発現レベルのわずかな変化も、高感度検出できる。これは、従来対数変換(log(ratio))では、微量な遺伝子発現比に対する検出力が低かったのに対して、arctan変換量を適用するためである。これにより、わずかな発現比の変化であっても、検出力の向上を図ることができる。

40

【0153】

さらに、対応分析による解析を行うため、解析結果が認識しやすい。また、対応分析から得られたスコアを用いることによって、既知の発現プロファイルから、より正確な表現型の予測を行うこともできる。

【0154】

さらに、従来、解析対象とされていなかった、発現比が比較的小さい発現プロファイルデータも表現型に關与するデータと推定することが可能である。それゆえ、表現型に關与

50

する新しい遺伝子やタンパク質を検出できる。

【0155】

すなわち、発現比の変化に有意性があるにもかかわらず、従来は見落とされていた等発現付近のデータであっても、本発明では確実に表現型に關与するものとして検出できる。すなわち、本発明によって得られたデータは、信頼性の高いデータである。それゆえ、従来は見落とされていた、有意性のある遺伝子やタンパク質を検出できる可能性が極めて高い。

【0156】

また、補足データを用いることによって、表現型に關連しないデータは除去され、さらに、ハウスキーピング遺伝子など常に発現量が一定のデータの推定も可能であるため、表現型に關連するデータのみを正確に推定できる。

【0157】

このように本発明によれば、膨大な量のマイクロアレイデータから、表現型に關連する遺伝子またはタンパク質のみを、解析結果を示すグラフによって容易に判断することができる。さらに、この解析は、特別なコンピュータを必要とせず、標準的なコンピュータによって、大規模のデータの解析を短時間かつ効率的に行うことができるという利点もある。したがって、従来のクラスター解析や主成分分析では困難であった、大規模データに対する迅速かつ明確な解析結果をもたらす斬新かつ強力なツールとして、極めて有用である。それゆえ、目的とする新規機能遺伝子の発見、ひいては、新薬の開発や、各種病態の遺伝子レベルでの解析などへの利用が期待される。

【0158】

なお、本実施形態では、対応分析の結果を3次元で示しているため、UDL、UDRの形状が、円柱や球になっている。しかしながら、UDL、UDRなどのすべての基準・計算は、3次元で有意距離などを算出するのではなく、UDL、UDRなどのもとのn次元で定義されるものである。

【0159】

(3) 本発明の発現プロファイル解析システムおよび解析方法の利用

また、本実施形態にかかる遺伝子発現プロファイル解析方法は、表現型が既知サンプルの発現プロファイルデータを用いることによって、表現型が未知サンプルの表現型の予測に利用可能である。具体的には、対応分析によって得られるn次元スコアから、未知サンプルから各既知サンプルまでの距離 D_{ij} は、計算可能である。ここで、例えば、表現型が二種類の場合にはiは1または2であり、それぞれ表現型AおよびBを示す。また、 j_k (kは既知サンプル数)である。例えば、表現型が未知のあるサンプルに着目したとき、 D_{11} は表現型がAである1番目のサンプルとの距離をあらわす。

【0160】

未知サンプルと既知サンプルとの距離が最小である場合、互いの遺伝子発現プロファイルは、最も類似している。しかしながら、同じプロファイルが、常に、量的形質における同一表現型をもたらすとは限らない。それゆえ、ここでは、全既知サンプルの距離と表現型を、全未知サンプルの表現型の予測に用いる。

【0161】

例えば、表現型が未知のあるサンプルに着目し、そのサンプルと最も近傍に位置する表現型が既知のサンプルが見つかるとする。未知サンプルと、近傍に位置する既知サンプル、すなわち、遺伝子発現プロファイルとは類似している。それゆえ、互いの表現型も似ていると推定できる。しかし、このように結論できるのは、環境による影響が少ない場合である。表現型は、遺伝子と環境の両者によって変化する。例えば、一卵性双生児でも環境が違えば寿命も異なる。このため、遺伝子発現プロファイルが最も類似している既知の1サンプルだけから、未知サンプルの表現型の推定を行うことは危険を伴う。

【0162】

そこで、本発明を未知サンプルの表現型の予測に利用する場合、表現型が未知のサンプルに対して、表現型がAである全員との距離を算出する。同様に、表現型がBである全員

10

20

30

40

50

との距離も算出する。さらに、前述のような推定の危険を避けるために、未知サンプルに対して、最も近傍に位置する一人からではなく、全員の距離を用い一般化することによって、未知サンプルの表現型を予測することを特徴とする。

【0163】

具体的には、未知サンプルから、これら表現型A群とB群の距離を比較して、いずれの群に距離的に近いかを調べ、表現型が未知のサンプルの表現型を推定する。

【0164】

表現型の予測では、表現型A群とB群に対する相対的な距離を求める式、すなわち下記式(2)と(3)に、各距離の逆数の和を用いているという特徴がある。

【0165】

【数2】

$$\sum_j (1/D_{1j}) \quad \dots \dots (2)$$

【0166】

【数3】

$$\sum_j (1/D_{2j}) \quad \dots \dots (3)$$

【0167】

一般的な考えでは、逆数ではなく距離そのものの総和を用いる。すなわち、表現型A群に対する距離の総和とB群に対する距離の総和とを比較し、距離が小さい方を、表現型が未知のサンプルの表現型と推定すると考えられる。

【0168】

しかし、この方法では、はずれ値をもつサンプル(個体)では、距離を過大評価することになる。すなわち、何らかの原因で、遠くに離れたプロットがあれば、その1個体の存在によって、距離の総和は急に大きくなってしまふ。このようなはずれ値は通常の解析では無視される。

【0169】

本発明を利用して表現型の予測を行う場合、はずれ値の影響も抑えながら、対応分析では、より近傍に位置する個体を重視し、逆に、表現型が未知のサンプルから遠く離れたところに位置するサンプルは、遺伝子発現プロファイルは大きく異なるので、表現型の推定に利用はするものの、推定に当たっては軽視することによって、適切な予測が可能となる。そこで、表現型の予測において、距離の逆数を用いることによって、より近傍に位置する個体を重視し、遠いサンプルを軽視するための一種の重み付けを行う。

【0170】

これにより、サンプル間の距離が近いければ近いほど、距離の逆数は+ に近づき、遠いほど0に近づく。すなわち、この距離の逆数の和を取るこの方法を用いることで、より近傍に位置しているものほど、式(2)と式(3)によって得られる値を巨大化することができる。

【0171】

式(2)と式(3)から得られる値は、表現型が未知のサンプルが、相対的にB群よりもA群に近いほど、式(2)の方が大きくなる。また、その逆も成立する。

【0172】

すなわち、未知サンプルが表現型Bよりも表現型Aサンプルの大部分に相対的に近い場合、式(2)の値は、式(3)の値よりも大きくなる。これは、未知サンプルが表現型Aであることを示している。一方、式(2) < 式(3)の場合、予測表現型はBとなる。この計算は、母集団の大部分と著しく異なった外れ値が、ほとんど式(4)に影響しないように、距離ではなく、互いの距離を用いている。

【0173】

10

20

30

40

50

【数4】

$$\sum_j (1/D_{ij}) \quad \dots \dots (4)$$

【0174】

一方、未知サンプルに最も近いサンプルは、この量に影響を及ぼす。その結果、表現型の予測が、発現プロファイルから正確に行うことができる。

【0175】

このように、類似の遺伝子発現プロファイルを持つサンプルは、対応分析の結果、同様のn次元座標を持ち、プロットした際には近傍に位置する。したがって、サンプル間のn次元距離を算出し、近距離であるほど、サンプルは類似の遺伝子発現プロファイルを持っていると判定することが可能となる。

10

【0176】

表現型を支配する遺伝子を、UDRを用いて決定し、次に、アレイ上のすべての遺伝子ではなく、UDRで決定された遺伝子の発現プロファイルだけを用いて、再度、対応分析を行うことによって、その結果得られた各サンプルのn次元座標は、表現型を支配している遺伝子の発現量だけで決定される。それゆえ、形質に無関係な遺伝子を解析に用いることなく、サンプルの表現型を判定できる。

【0177】

ここで、環境による影響（食生活、年齢など）が形質の変化にさほど大きな影響を及ぼさないと仮定すると、表現型を支配している遺伝子群が同様の発現量を示しているサンプルは、同様の表現型になると予測される。表現型を支配している遺伝子群が同じ遺伝子発現量を示している個体間なら、当然、同様の表現型を示す。

20

【0178】

そこで、ある表現型を支配している遺伝子の発現量と表現型の関係をデータベースのように蓄積していくことで、新たな被験者（サンプル）がどのような表現型を示すかの予想に利用できる。

【0179】

例えば、表現型を支配している20の遺伝子の発現プロファイルが表現型のわかっている100人のサンプルで既知であるとすれば、100人のデータから、20の遺伝子の発現プロファイルがどのような表現型になっているかがわかることを意味する。この表現型が未知のサンプルを、表現型が既知の100人の座標値と比較して、n次元距離を求め、相対的にどの表現型のサンプルに近いかを判定する。この際、遺伝子発現プロファイルに近いほど類似の表現型となるので、表現型未知のサンプルの表現型が予測可能となる。

30

【0180】

また、この表現型が、例えば、計測に時間とコストがかかる；計測精度が低い；サンプルの表現型が数週間後や10年後にどうなるか；育種素材としていい父本、母本となりえるか；乳牛としての良否を子牛のときに知りたい；などといった場合などのように、計測困難もしくは不可能な表現型であるとする。この際、表現型がわからないサンプルが加わっても、そのサンプルの遺伝子発現プロファイルを計測し、対応分析を行うことによって、その表現型が分からないサンプルの座標が得られる。

40

【0181】

しかし、遺伝的に同一な一卵性双生児であっても、環境による表現型の変動が生じる。したがって、まったく同一の遺伝子発現プロファイルの個体（サンプル）が二個体あったとしても、まったく同じ血圧、身長にはなる可能性は低い。例えば、計測した日の体調や環境（食生活、生活リズム）にも左右される。それゆえ、最も近傍（類似の発現プロファイル）だけで表現型を予測するのは危険である。

【0182】

そこで、表現型が複数に大きく分類されるとき、表現型が未知の個体を、複数の表現型

50

に属するすべての個体群との距離を用いて、どの表現型群に相対的に近いかを見ることによって、表現型を推定する。たとえば、表現型が3つのA、B、C群に分かれ、それぞれ、サンプル数が10人、15人、13人であるとする、まず、表現型が未知のサンプルから表現型Aの10人に至るまでの、10個のn次元距離を算出する。表現型B、C群に対しても同様にして、n次元距離を算出する。そして、A、B、C群、それぞれにおいて、距離の逆数の和をとる。このA、B、C群で算出した値のうち最大であった群の表現型を、表現型未知の表現型と推定する。

【0183】

このように、本発明の発現プロファイル解析システムを使用して、膨大な発現プロファイルデータを網羅的に迅速かつ網羅的に解析し、その解析結果から、目的とする表現型に
10
関与する遺伝子またはタンパク質を容易に検出・推定・同定・予測することができる。目的とする表現型としては、例えば、ヒトを含む生物の生体内環境、例えば、分化、成長、老化、代謝、疾病の等が挙げられる。本発明により、これらのモニタリングや疾病発症可能性の予測、診断が可能となる。

【0184】

特に、ヒトの疾病の診断や治療法の選択においては、当該疾病の特徴を分子レベルで把握することが有用である。各種の疾病には多数の遺伝子やタンパク質の発現量の変動が起こっており（つまり発現比が異なる）、この変動はその原因、病状、個体差によってそのパターンが異なる。すなわち、いくつかの遺伝子やタンパク質の発現プロファイルデータは、各個体における疾病の性格や病状を反映しており、その解析によって診断、治療に有用なデータを抽出することが可能である。この有用なデータとは、例えば、疾病の名称、
20
タイプ、原因、進行状況、予後、余命、薬剤に対する感受性やその副作用、発症、再発、転移の可能性等がある。本発明により、これらの表現型に関与する有用なデータの推定・分類・予測法を効率よく、かつ高い精度で得ることができる。

【0185】

例えば、子供では発現しないが、大人で発現するために疾患となる場合、その疾患に関与する遺伝子やタンパク質を検出することによって、疾患の発症可能性を予測することが可能となる。それゆえ、本発明は、テーラーメイド医療への応用も期待される。

【実施例1】

【0186】

本実施例では、上記遺伝子発現プロファイル解析システムを用いて、公知のヒト乳癌患者のマイクロアレイデータを用い（L. J. Veer et al., Nature. 415, 530 (2002).）、
30
遺伝子発現プロファイル解析を行った。すなわち、本実施例では、マイクロアレイデータから、表現型A（癌）またはB（非癌）に関与する遺伝子を推定した。なお、上記マイクロアレイデータは、発現プロファイルデータに（ $\log_{10}(\text{ratio})$ ）が用いられているため、 $\arctan(1/\text{ratio})$ 値に変換して解析を行った。

【0187】

本実施例で用いたマイクロアレイデータは、115サンプルに対する、24024個の利用可能な遺伝子発現比を有している。アレイ上の24024個の遺伝子は、以下の4つのカテゴリーに分類できる。すなわち、（a）癌、または、非癌条件において、特異的に
40
発現した遺伝子；（b）癌条件と非癌条件とを比較した場合に、発現量が増減し、かつ、癌に関与する遺伝子；（c）各条件でランダムに発現量が増減するが、癌に関与しない遺伝子；（d）全サンプルほぼ同じ発現レベルを示すハウスキーピング遺伝子、すなわち、母集団の表現型の区別に直接影響しない遺伝子。これらのカテゴリーのうち、（a）および（b）が、疾患に関与する遺伝子を含んでいる。

【0188】

つまり、本実施例では「癌に関与する遺伝子」として、癌特異的に発現する遺伝子、非癌特異的に発現する遺伝子、癌において発現量が変化する遺伝子、および、非癌において発現量が変化する遺伝子を推定することになる。この解析によって、新しい癌関与遺伝子の発見につながり、新規医薬品の開発に役立つ。
50

【 0 1 8 9 】

本実施例では、公知マイクロアレイデータに、24024個の遺伝子のカテゴリー分類を容易にするために、第1～第3補足データ(Supplement1-3)を加えた。

【 0 1 9 0 】

第1補足データの発現比は、全癌サンプルではゼロである一方、全非癌サンプルでは最大発現比を示す。第2補足データの発現比は、第1補足データのパターンと逆の遺伝子発現パターンを有している。癌サンプルまたは非癌サンプルで特に発現するこれら2つの補足遺伝子は、癌関連遺伝子を発見するために有用である。

【 0 1 9 1 】

第3補足データは、全サンプルについて同じ発現比(1倍付近、等発現)であり、ハウスキーピング遺伝子を同定するために用いる

パーソナルコンピュータによる対応分析計算では、全24024遺伝子、115サンプルは、数分間で、7次元スコア(座標)となった。

【 0 1 9 2 】

前述の第1および第2補足データは、異なる7次元スコアを有している。対応分析の理論によると、疾患によって誘導または抑制された遺伝子の7次元スコアは、前記2つの遺伝子間で、次第に、直線的に変化する。ここで、この7次元の直線を「UDL」とする。しかしながら、その他の生物学的解析においても、誘導または抑制された遺伝子のスコアを、UDLから統計的に分類している。

【 0 1 9 3 】

本実施例では、UDLから、有意な領域を規定するために、7次元における有意なカイ二乗距離を用いた。ある遺伝子とUDLの7次元距離とが、有意距離以上である場合、その遺伝子は、深く癌に関係していないと推定した。以下では、有意領域を、「UDR」と称する。

【 0 1 9 4 】

これにより、UDRの外部に位置した23928個の遺伝子を、癌に関与しない遺伝子と推定できた。

【 0 1 9 5 】

図6(a)は、本発明者が確立したクリックابلビューアー-e-GRED(clickable viewer e-GRED)による、全結果の3次元の部分空間を示している。7次元で規定したUDRは、3次元では、円筒形状を示した。第1～第3主軸(Factor1~3)までの累積寄与率は、62.5%であり、e-GRED図は、表現型に関与する遺伝子と、表現型に関与しない遺伝子とを有意に示した。

【 0 1 9 6 】

UDRの内部には、ハウスキーピング遺伝子を含む可能性がある(カテゴリー(d))。対応分析では、ハウスキーピング遺伝子は、上記第3補足遺伝子の位置の周囲に集まる。そこで、本実施例では、第3補足遺伝子の位置からの有意な7次元のカイ二乗距離を算出して、ハウスキーピング遺伝子を推定した。

【 0 1 9 7 】

図6(b)に示すように、この有意領域は、3次元グラフにおいて、球状となる。したがって、この球の内部に位置するサンプルは、ハウスキーピング遺伝子として推定できる。これにより、UDRの内部に位置した遺伝子から、ハウスキーピング遺伝子を排除し、癌に関与する遺伝子のみを検出した。

【 0 1 9 8 】

なお、図6において、面内の線は軸を示している。また、図6(a)の円柱はUDRであり、その中心線はUDLである。23928の癌に関与しない遺伝子は、UDRの外側にある。

【 0 1 9 9 】

また、図6(b)では、96の遺伝子がUDRの内部にある。そのうち、67の遺伝子は、球内に存在し、ハウスキーピング遺伝子を示している。残りの29遺伝子のうち、1

10

20

30

40

50

5の赤球と14の緑球は、それぞれ、癌によって統計的にアップレギュレーションまたはダウンレギュレーションされたことを示している。

【0200】

このようにして、本実施例では、これらの有意領域を用いて、29個の癌関与遺伝子を同定した。

(比較例1)

【0201】

実施例1の対応分析による推定結果を確認するために、上記29個の癌関与遺伝子の発現プロファイルを有する115サンプルの、従来の階層クラスタリング解析を行った。この解析では、遺伝子とサンプルのいずれも、2つのクラスターを作成した。(図7)。すなわち、左右のサンプルクラスターにある、53個の非癌サンプルのうちの39サンプル、および、62個の癌サンプルのうちの40サンプルを、それぞれ分類した。(図7A・D)。この結果は、全体として、左側のクラスターは非癌サンプルの遺伝子発現プロファイルを示しており、右側のクラスターは癌サンプルの遺伝子発現プロファイルを示している。図7中の直交する線に一致するクラスターにおける14個の遺伝子は、非癌サンプルで誘導(発現が増加)した。一方、その線より下のクラスターにおける15個の遺伝子は、癌サンプルで誘導(発現が増加)した。(図7B・C)。

10

【0202】

なお、図7は、115サンプル中、癌に関与すると推定される29遺伝子の2次元クラスター解析の発現プロファイルである。図7中、Aはサンプルのクラスタリング。Bは遺伝子のクラスタリング。Cは各行と列は、それぞれ、サンプルと遺伝子とを示している。図中の濃淡は、それぞれ、発現が誘導または抑制されたことを示している。縦方向および横方向の垂線は、遺伝子およびサンプルの2つのクラスターにおける再分類を示している。Dは、黒は非癌サンプルを、白は癌サンプルを示している。

20

【0203】

Bに示す遺伝子のクラスターは大きく二つのサブクラスターに分類されている。このサブクラスターの遺伝子は、それぞれ、14個の遺伝子と、15個の遺伝子である。これらの遺伝子群は、対応分析の結果、第1主軸に沿って正負に分かれており、両解析結果は完全に一致している。

【0204】

前述のように、第1主軸が最大の寄与率を有し、第2主軸、第3主軸の順で寄与率が低くなる。したがって、第1主軸に沿って正負に分かれているということは、表現型を決定する遺伝子が、表現型を説明する複数の主軸のうち、第1主軸だけで説明が可能であったことを意味する。すなわち、第1主軸を見てガン側で強く発現する遺伝子群と弱く発現する遺伝子群が、それぞれ、正負にきれいに分離できた。なお、第1主軸だけできれいに分かれられない(寄与率が低く説明しきれない)ときは、第2主軸以降の正負もみればよい。

30

【実施例2】

【0205】

(予測可能性への応用例)

本実施例では、実施例1と同じデータを用いて、表現型の予測可能性を検討した。

40

【0206】

図8は、29個の癌関連遺伝子の発現プロファイルを有する115のサンプルのe-GREDを示している。癌サンプルの大部分は、逆に、第1軸(Factor1)に沿ってプラススコアで位置している。e-GREDによる不完全なサンプル分類は、累積寄与率が低いために生じる(66.4%)。しかしながら、この結果は、量的形質が制御されている遺伝子発現プロファイルのみが、分類を適切に判断できていないという可能性がある。遺伝子発現プロファイルが表現型を正確に予測することができる文献では(L. J. Veer et al., Nature. 415, 530 (2002).) (予測可能性約83%)、その結果は、母集団のみから得たものであった。その予測可能性は、使用する母集団によって変化する。遺伝子発現プロファイルデータのみから明らかにされる表現型分類の限度(範囲)は、最も重要な解決課題の1つで

50

ある。

【0207】

そこで、10000回のモンテカルロシミュレーションを、予測可能性の分布を説明するために行った。つまり、モンテカルロシミュレーションでは、各回とも、この115サンプルを二つのグループ(95個の管理サンプルと25個の非管理サンプル)にランダムに分割し、25個の表現型を何人まで正確に予測できるかを%で表した。これは、1回の試行だけでは予測精度を正しく評価できないため、数回の試行を行うことによって、試行ごとに予測精度の変動をみるためである。したがって、10000回行い、10000個の予測精度を算出した。

【0208】

管理サンプルの表現型は、全て分かっている。しかし、非管理サンプルの表現型は、未知である。1つの非管理サンプルから各管理サンプルまでの7次元距離を計算した。非管理サンプルが、非癌の管理サンプルに最も近い場合、非管理サンプルの表現型は、非癌と予測する(逆の場合も同様)。20個の非管理サンプルから、各回の予測可能性を導き出した。10000回の予測可能性の割合は、約70%(sd9.5%)であった。その範囲(限度)は、25~100%であり、8955個の予測可能性は、60%を越えた。

【産業上の利用可能性】

【0209】

本発明の発現プロファイル解析システムおよび解析方法は、マイクロアレイなどによって得られる膨大な数のデータを網羅的に解析するアプローチであり、表現型の関連および予測について量的形質を有する関連遺伝子を発見するためのハイスループット解析に利用できる。

【0210】

本発明で用いる対数変換指標は、従来の指標よりも、発現比が0.1~10倍において特に著しく変化する。これにより、量的形質が制御されているわずかな発現比の変動を、高感度検出することによって生じる。

【0211】

それゆえ、そのような遺伝子やタンパク質は、新薬のターゲットとして利用できる可能性があるので、新薬の開発など、医療分野での応用が期待される。また、生理活性物質の探索や薬物代謝の研究など、生物学、分子生物学、医学、薬学領域における基礎研究に留まらず、動物の育種や、テラーメード医療等に広くこの発明を利用することも可能である。

【図面の簡単な説明】

【0212】

【図1】本発明の実施の一形態にかかる遺伝子発現プロファイル解析システムの要部構成を示すブロック図である。

【図2】図1の遺伝子発現プロファイル解析システムの主な動作を示すフローチャートである。

【図3】本発明の対数変換値と従来の対数変換値とを、発現比に対してプロットしたグラフである。

【図4】本発明の実施の一形態にかかる遺伝子発現プロファイル解析の説明に用いたデータである。

【図5】図4のデータの解析結果を示すグラフであり、図5(a)は本発明の対応分析による解析結果であり、図5(b)は従来の主成分分析による解析結果である。

【図6】実施例1の解析結果を示す図であり、図6(a)は全解析結果であり、図6(b)は図6(a)の主要部のみを示す図である。

【図7】従来のクラスタリング解析結果を示す図である。

【図8】図6(b)のデータの主成分分析による解析結果を示す図である。

【図9】本発明の実施の一形態にかかる遺伝子発現プロファイル解析システムにおける、解析の手段として「R」を用いた解析プログラムである。

10

20

30

40

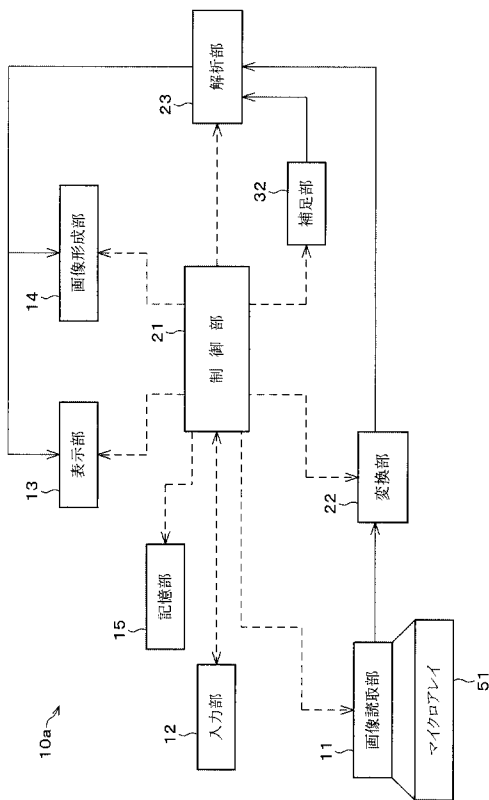
50

【符号の説明】

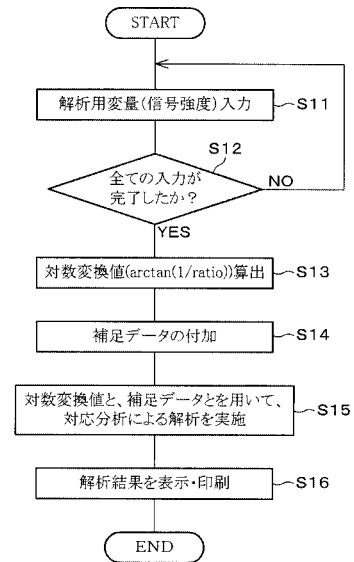
【0213】

- 10 a 解析システム（発現プロファイル解析システム）
- 22 変換部（変換手段）
- 23 解析部（解析手段）
- 32 補足部

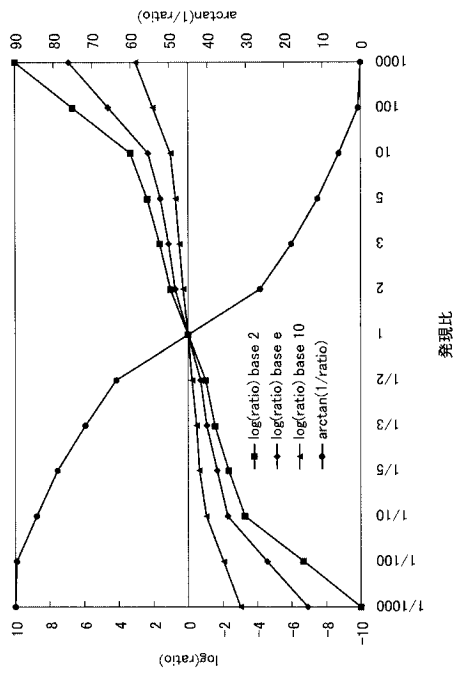
【図1】



【図2】



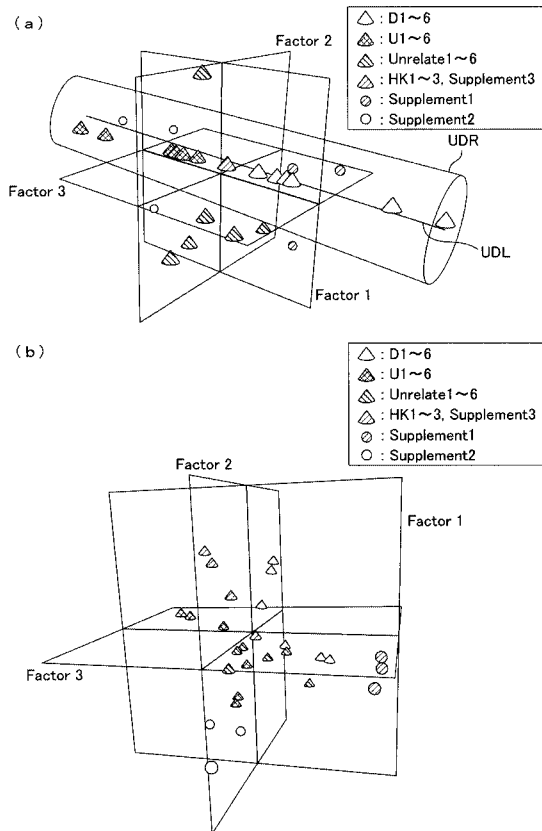
【 図 3 】



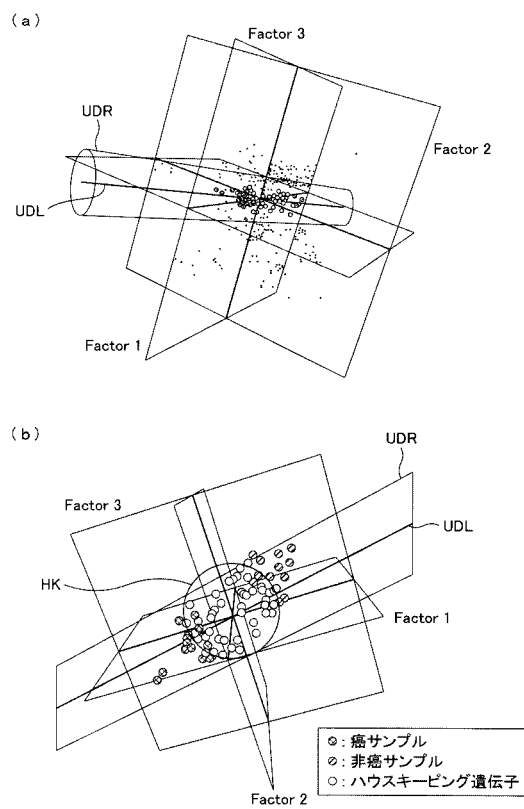
【 図 4 】

Gene	Sample					
	A1	A2	A3	B1	B2	B3
D1	0.01	0.01	0.01	1.00	1.00	1.00
D2	0.10	0.10	0.10	1.00	1.00	1.00
D3	0.50	0.50	0.50	1.00	1.00	1.00
D4	1.00	1.00	1.00	2.00	2.00	2.00
D5	1.00	1.00	1.00	10.00	10.00	10.00
D6	1.00	1.00	1.00	100.00	100.00	100.00
U1	2.00	2.00	2.00	1.00	1.00	1.00
U2	10.00	10.00	10.00	1.00	1.00	1.00
U3	100.00	100.00	100.00	1.00	1.00	1.00
U4	1.00	1.00	1.00	0.01	0.01	0.01
U5	1.00	1.00	1.00	0.10	0.10	0.10
U6	1.00	1.00	1.00	0.50	0.50	0.50
Unrelate1	0.50	10.00	0.50	10.00	0.50	0.20
Unrelate2	2.00	10.00	1.00	0.10	0.10	0.20
Unrelate3	0.10	0.00	0.50	0.10	10.00	0.10
Unrelate4	0.20	1.00	10.00	1.00	2.00	0.50
Unrelate5	0.50	0.50	0.50	1.00	10.00	0.20
Unrelate6	0.20	1.00	10.00	2.00	0.50	0.10
HK1	10.00	10.00	10.00	10.00	10.00	10.00
HK2	5.00	5.00	5.00	5.00	5.00	5.00
HK3	2.00	2.00	2.00	2.00	2.00	2.00
Supplement1	100.00	100.00	100.00	0.00	0.00	0.00
Supplement2	0.00	0.00	0.00	100.00	100.00	100.00
Supplement3	1.00	1.00	1.00	1.00	1.00	1.00

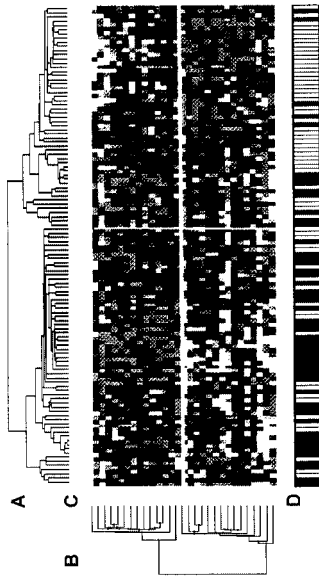
【 図 5 】



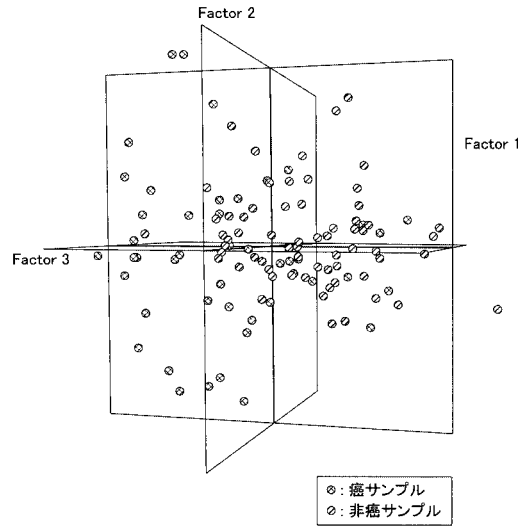
【 図 6 】



【 図 7 】



【 図 8 】



【 図 9 】

```

R: Copyright 2003, The R Development Core Team
Version 1.6.2 (2003-01-10)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> # The character "#" at first column means the line is comment.
> # load library multiv
> library(multiv)
> # read the example data (Table S1)
> Data <- as.matrix(read.table("example.txt"))
> # convert expression ratios to new indices
> In <- atan(1/Data)/pi*180
> # perform CA and write result to text files
> CAOUT <- ca(In)
> write.table(CAOUT$proj, "CArow.txt", sep="\t")
> write.table(CAOUT$proj, "CAcolumn.txt", sep="\t")
> write.table(CAOUT$vals, "CAevals.txt", sep="\t")
> # perform PCA and write result to text files
> PCAOUT <- pca(In)
> write.table(PCAOUT$proj, "PCArow.txt", sep="\t")
> write.table(PCAOUT$proj, "PCAcolumn.txt", sep="\t")
> write.table(PCAOUT$vals, "PCAEvals.txt", sep="\t")
> █

```

フロントページの続き

- (72)発明者 佐藤 和広
岡山県倉敷市有城1169-40
- (72)発明者 武田 和義
岡山県倉敷市有城1169-71

審査官 三原 健治

(56)参考文献 特開2007-199948(JP,A)

(58)調査した分野(Int.Cl., DB名)

C12N 15/00-15/90

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

JSTPlus/JMEDPlus/JST7580(JDreamII)

G-Search