

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6565801号
(P6565801)

(45) 発行日 令和1年8月28日(2019.8.28)

(24) 登録日 令和1年8月9日(2019.8.9)

(51) Int.Cl. F I
GO 1 N 27/62 (2006.01) GO 1 N 27/62 D

請求項の数 8 (全 16 頁)

<p>(21) 出願番号 特願2016-124083 (P2016-124083) (22) 出願日 平成28年6月23日 (2016. 6. 23) (65) 公開番号 特開2017-227542 (P2017-227542A) (43) 公開日 平成29年12月28日 (2017.12.28) 審査請求日 平成30年9月26日 (2018. 9. 26)</p>	<p>(73) 特許権者 000001993 株式会社島津製作所 京都府京都市中京区西ノ京桑原町1番地 (74) 代理人 110001069 特許業務法人京都国際特許事務所 (72) 発明者 山田 賢志 京都市中京区西ノ京桑原町1番地 株式会 社島津製作所内 審査官 吉田 将志</p>
--	--

最終頁に続く

(54) 【発明の名称】 質量分析データ処理装置、質量分析装置、質量分析データ処理方法、及び質量分析データ処理用プログラム

(57) 【特許請求の範囲】

【請求項1】

複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサンプルに対するマススペクトルデータを処理する質量分析データ処理装置であって、

a) 与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成するピークマトリクス作成部と、

b) 前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つのピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類するレベル分け部と、

c) 前記ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別するレベル識別部と、

d) 前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得するマトリクス修正部と、

を備え、前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴とする質量分析データ処理装置。

10

20

【請求項 2】

請求項 1 に記載の質量分析データ処理装置であって、前記レベル識別部は、

c1) 前記ピークマトリクスの行毎に統計的仮説検定における p 値を計算する p 値計算部と、

c2) 強度値レベル毎に、一つの強度値レベルに分類された複数の行に対応する p 値の値に関する度数分布を求める度数分布取得部と、

c3) 強度値レベル毎に、p 値の値に関する度数分布からその度数の一様性を判定し、度数の一様性が高い強度値レベルを解析不適レベルであると判断する判定部と、
を含むことを特徴とする質量分析データ処理装置。

【請求項 3】

請求項 1 又は 2 に記載の質量分析データ処理装置であって、

前記フィルタリング済みピークマトリクスに基づいてグループ間の差異解析を実行し、該差異解析において有意差がある行に対応するピーク又は質量電荷比を抽出する差異解析部、をさらに備えることを特徴とする質量分析データ処理装置。

【請求項 4】

請求項 1 ~ 3 のいずれか 1 項に記載の質量分析データ処理装置を含む質量分析装置であって、

サンプルに対し質量分析を行うことでマススペクトルデータを取得する質量分析部と、複数のグループのいずれかに属する複数のサンプルに対し前記質量分析部で質量分析を実行することで得られたマススペクトルデータを格納するデータ記憶部と、

をさらに備え、前記データ記憶部に格納されているマススペクトルデータを前記質量分析データ処理装置のピークマトリクス作成部に与えることを特徴とする質量分析装置。

【請求項 5】

複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサンプルに対するマススペクトルデータを処理する質量分析データ処理方法であって、

a) 与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成し、

b) 前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つのピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類し、

c) 前記ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別し、

d) 前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得し、

前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴とする質量分析データ処理方法。

【請求項 6】

請求項 5 に記載の質量分析データ処理方法であって、

前記解析不適レベルを識別するために、前記ピークマトリクスの行毎に統計的仮説検定における p 値を計算し、強度値レベル毎に、一つの強度値レベルに分類された複数の行に対応する p 値の値に関する度数分布を求め、強度値レベル毎に、p 値の値に関する度数分布からその度数の一様性を判定し、度数の一様性が高い強度値レベルを解析不適レベルであると判断することを特徴とする質量分析データ処理方法。

【請求項 7】

複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサン

10

20

30

40

50

ルに対するマススペクトルデータを処理するべくコンピュータ上で動作する質量分析データ処理用プログラムであって、

a)与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成するピークマトリクス作成ステップと、

b)前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つのピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類するレベル分けステップと、

c)前記ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別するレベル識別ステップと、

d)前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得するマトリクス修正ステップと、

をコンピュータに実行させ、前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴とする質量分析データ処理用プログラム。

【請求項 8】

請求項 7 に記載の質量分析データ処理用プログラムであって、前記レベル識別ステップとして、

c1)前記ピークマトリクスの行毎に統計的仮説検定における p 値を計算する p 値計算ステップと、

c2)強度値レベル毎に、一つの強度値レベルに分類された複数の行に対応する p 値の値に関する度数分布を求める度数分布取得ステップと、

c3)強度値レベル毎に、p 値の値に関する度数分布からその度数の一様性を判定し、度数の一様性が高い強度値レベルを解析不適レベルであると判断する判定ステップと、

をコンピュータに実行させるようにしたことを特徴とする質量分析データ処理用プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、質量分析により得られたマススペクトルデータを処理する質量分析データ処理方法、該データ処理方法を実施する質量分析データ処理装置、及びそのためのコンピュータプログラム、さらには、上記質量分析データ処理装置を備えた質量分析装置に関し、さらに詳しくは、複数のサンプルグループの間での差異解析を行うのに好適な質量分析データ処理装置、質量分析装置、質量分析データ処理方法、及び質量分析データ処理用プログラムに関する。

【背景技術】

【0002】

近年、特定の疾病や疾患の早期診断や治療効果の確認などのために、質量分析法を利用したバイオマーカーの解析の研究が進められ、一部は実用に供されている。例えば複数の健常者から採取した血液、尿等の生体試料には全く又は殆ど存在せず、ガンなどの特定の疾患を罹患している複数の患者から採取した血液、尿等の生体試料に明確に存在している物質が見つかれば、その物質はその疾患における有力なバイオマーカー候補であるといえる。一般にこうしたバイオマーカーの探索は、二又はそれ以上の複数のグループ（例えば健常者グループと患者グループ）にそれぞれ由来する多数のサンプルを質量分析装置により測定し、それにより得られたデータについてグループ間での差異解析を行うことで行われる。

【0003】

バイオマーカー（又はより一般的なマーカー）を探索することを目的とした従来の一般的な質量分析データの差異解析の手順は以下のとおりである。ここでは、グループの総数を N_G 、サンプルの総数を N_S であるとする。

【ステップ A 1】各グループに属するサンプルに対するマススペクトルをそれぞれ取得し、各マススペクトルにおいてピーク強度値が所定の閾値 I を超えているピークをノイズレベルを超えているピークとみなして検出する。そして、マススペクトル毎に、検出されたピークの質量電荷比 (m/z) 値とピーク強度値とをまとめたピークリストを作成する。作成されるピークリストの総数はサンプルの総数と同じ N_S であり、それらは N_G 個のグループに分類可能である。

【0004】

【ステップ A 2】上記ステップ A 1 で得られたピークリストを列ベクトルと考え、同じ質量電荷比に対するピーク強度値が同一行となるように、ピーク強度値を配置した行列（ピークマトリクス） M_p を作成する。或る一つのマススペクトルにおいて検出されたピークの質量電荷比に他のマススペクトルではピークが存在しない（ピークリスト中に対応するピークがない）場合には、他のマススペクトルにおけるそのピークのピーク強度値をゼロとすればよい。図 4 (a) はピークマトリクスの概念図である。ここでは、ピークマトリクスの横方向（行方向）にサンプル情報（例えばサンプル番号）、縦方向（列方向）に質量電荷比値をとり、行列の要素がピーク強度値である。このとき、ピークマトリクス M_p の列数はサンプル数 N_S に等しく、ピークマトリクス M_p の行数は、サンプル全体を通して検出されたピークの総数（重複を省く） N_p に等しい。

【0005】

【ステップ A 3】ステップ A 2 で作成されたピークマトリクス M_p の各行に対し、グループ間の差異の有無を調べるために単変量解析（一般には t 検定、 U 検定、分散分析 (ANOVA) 等の統計的仮説検定) を実施し、観察された差異の統計的信頼性を示す p 値 (p -value) を行毎に計算する。

【ステップ A 4】各行の p 値を予め定めた有意水準（例えば t 検定では $\alpha = 0.05$ がしばしば用いられる）と比較することにより、グループ間で有意差のある行つまりはピーク（又は質量電荷比）を抽出する。この抽出されたピークに対応する成分がバイオマーカー候補である。

【0006】

ピークマトリクス M_p の行数 N_p が 1（ピークが一つのみである）の場合には、ステップ A 3 において求めた p 値を単に有意水準 α と比較することによって、グループ間で有意差があるか否かという検定を行うことができる。しかしながら、一般的にマススペクトルには多数のピークが存在するため $N_p \gg 1$ である。その場合、有意水準 α での検定を行毎に行うと、実際にはグループ間に差異がないにも拘わらず、偶然に「有意差あり」と判定されるピークが含まれる可能性が高くなるという問題がある。こうした現象は統計的仮説検定において多重比較を行う際の問題としてよく知られており、多重検定（又は検定の多重性）と呼ばれている。

【0007】

上記多重検定の問題を回避して検定の正確性を高めるためには、有意水準 α の定義を見直し、以下の A 又は B のいずれかの値が所定値以下になるように検定を行う必要がある。

(A) Familywise Error Rate (FWER) : 実際には有意差がないピークのうち、一つでも「有意差あり」と誤判定される確率

(B) False Discovery Rate (FDR) : 有意差ありと判定されたピークのうち、実際にはグループ間での差異がないピークの比率

【0008】

上記 (A) の FWER を所定値以下にするという方針に従った手法としてよく知られているのは Bonferroni 法であり、これは簡単に言うと、それぞれの検定の有意水準を α / N_p に変更する方法である。しかしながら、この方法は、検定の繰り返し数（つまりは N_p ）が増加するに従い検出力が低くなり、実際には差異があっても「有意差なし」と

10

20

30

40

50

誤って判定してしまう可能性（一般に偽陰性又は エラーと呼ばれる）が高くなるという問題がある。即ち、上記方法では、実際には差異がないにも拘わらず「有意差あり」と誤って判定してしまう可能性（一般に偽陽性又は エラーと呼ばれる）が生じることを厳密に抑えすぎるために エラーが高くなる。Bonferroni法よりも有意水準をやや緩めた方法としてHolm法もあるが、この方法でも検定の繰り返し数が多い場合に エラーが高いのはBonferroni法と同様である。

【0009】

上記理由から、従来の多重比較検定では、上記（B）のFDRを調整する手法が一般的である。FDRを所定値以下に抑える検定の手順としては、非特許文献1に記載のBenjamini & Hochberg法（BH法）やそれをベースとして改良した方法が知られている。これら方法は、エラーを或る程度許容しつつ エラーが生じる可能性を抑えるものである。

10

【0010】

しかしながら、非特許文献1等に記載の方法によってFDRを有意水準以下に抑えるためには、検定の繰り返し数が事前に分かっている必要がある。上述したように、マススペクトルデータを用いた差異解析では、検定の繰り返し数はピークマトリクスの行数 N_p に相当するが、或るデータに基づいて作成されるピークマトリクスの行数は一義的に決まっているわけではなく恣意性が存在する。何故なら、上述の差異解析手順のステップA1において閾値Iを小さくすると検出されるピーク数が増え、ピークマトリクスの行数が増大するためである。マススペクトル上でピーク強度が低いピークとノイズとを分ける閾値Iを自動的に決めることは難しい。そのため、通常、閾値Iは分析者が設定する機会が多いが、この閾値Iの増減によって検出されるピークの数に変動することになる。

20

【0011】

非特許文献1に記載のBH法による多重比較検定では、ピークマトリクスの行数 N_p が多くなるほど、同じピークであっても「有意差なし」と判定され易くなる。そのため、閾値Iが変化することによってピークマトリクスの行数 N_p が変化すると、各ピークに対する検定結果も変化してしまうおそれがある。こうした検定結果の変動は、ピーク強度が閾値I付近であるピークのみならず、ピーク強度が閾値Iよりも十分に大きなピークにも及ぶ。そのため、閾値I（つまりはノイズレベル）の決め方によっては、十分に明瞭なピークに対する検定結果にまで影響が及ぶことになり、検定結果つまりは差異解析結果の信頼性を損ねることになる。

30

【先行技術文献】

【非特許文献】

【0012】

【非特許文献1】ベンジャミン（Y. Benjamin）、ほか、「コントローリング・ザ・フォールス・ディスカバリー・レート：ア・プラクティカル・アンド・パワフル・アプローチ・トゥ・マルチプル・テストング（Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing）」、ジャーナル・オブ・ザ・ロイヤル・スタティカル・ソサイエティ：シリーズ B（Journal of the Royal Statistical Society: Series B）、1995年、Vol. 57、No. 1、pp. 289-300

【発明の概要】

40

【発明が解決しようとする課題】

【0013】

本発明は上記課題を解決するためになされたものであり、その目的とするところは、例えばそれぞれが複数のサンプルを含む複数のグループを識別するマーカーを見つけるために各サンプルに対して得られたマススペクトルデータに基づく差異解析を行う際に、マススペクトルからピークを検出するための信号強度の閾値、つまりはノイズレベルの設定の影響を受けない正確な差異解析を行うことができる質量分析データ処理装置、質量分析装置、質量分析データ処理方法及び質量分析データ処理用プログラムを提供することである。

【課題を解決するための手段】

50

【0014】

上記課題を解決するために成された本発明に係る質量分析データ処理装置は、複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサンプルに対するマススペクトルデータを処理するデータ処理装置であって、

a)与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成するピークマトリクス作成部と、

b)前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つのピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類するレベル分け部と、

c)前記ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別するレベル識別部と、

d)前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得するマトリクス修正部と、

を備え、前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴としている。

【0015】

また上記課題を解決するために成された本発明に係る質量分析データ処理方法は、複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサンプルに対するマススペクトルデータを処理する質量分析データ処理方法であって、

a)与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成し、

b)前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つのピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類し、

c)前記ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別し、

d)前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得し、

前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴としている。

【0016】

また上記課題を解決するために成された本発明に係る質量分析データ処理用プログラムは、複数のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータに基づいてグループ間の差異解析を行うために、該複数のサンプルに対するマススペクトルデータを処理するべくコンピュータ上で動作する質量分析データ処理用プログラムであって、

a)与えられた複数のサンプルに対するマススペクトルからそれぞれピークを検出し、そのピークの質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素として配置したピークマトリクスを作成するピークマトリクス作成ステップと、

b)前記ピークマトリクスにおいて、行毎にそれぞれ、一つの行に含まれる代表的な一つ

10

20

30

40

50

のピーク強度値又は複数のピーク強度値に基づいて算出された代表値を利用してその行を複数段階の強度値レベルに分類するレベル分けステップと、

c)前記ピークマトリクス of 行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値を算出し、同じ強度値レベルに分類された複数の行に対応する前記指標値の度数分布に基づいて、前記複数段階の強度値レベルの中でグループ間の差異解析が不可能である又は困難である解析不適レベルを識別するレベル識別ステップと、

d)前記ピークマトリクスから前記解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得するマトリクス修正ステップと、

をコンピュータに実行させ、前記フィルタリング済みピークマトリクスをグループ間の差異解析に供するようにしたことを特徴としている。

10

【0017】

なお、上記ピークマトリクスにおいて行と列とは便宜的に定められたものであるから、行と列を入れ替えることができることは明らかである。即ち、上記各発明において「行」を「列」に置き換えても実質的に同じであることは明らかである。

【0018】

本発明に係る質量分析データ処理装置における処理対象であるデータは、二以上のグループのいずれかに属する複数のサンプルに対し質量分析を行うことで得られたマススペクトルデータである。各グループはそれぞれ複数の、通常は多数のサンプルを含む。また、マススペクトルはMRM (Multiple Reaction Monitoring) 測定における定量用のMSⁿ スペクトルでもよい。

20

【0019】

本発明に係る質量分析データ処理方法を実施するための本発明に係る質量分析データ処理装置において、ピークマトリクス作成部は、上述したようなマススペクトルデータが与えられると、各マススペクトルからそれぞれ所定の基準に従ってピークを検出する。所定の基準は特に限定されず、一般に用いられているピーク検出方法を用いることができる。例えば、マススペクトル上で観測されるピーク状の信号のうち、信号強度が所定値以上であるものをピークとして検出し、信号強度が所定値未満であるものはノイズであるとみなせばよい。そうしてマススペクトル毎に検出したピークの信号強度値つまりピーク強度値と質量電荷比値を取得し、質量電荷比値を行方向に、サンプルを識別する情報を列方向に割り当て、ピーク強度値を要素としたピークマトリクスを作成する。ピークマトリクスにおける一つの行は、或る一つの質量電荷比における全てのサンプルに対するピーク強度値である。

30

【0020】

レベル分け部は、作成されたピークマトリクスにおいて、行毎にそれぞれ、例えば一つの行に含まれる複数のピーク強度値の中で最大値を代表的な一つのピーク強度値として求め、この値に基づいてその行を複数段階の強度値レベルのいずれかに分類する。強度値レベルの値の範囲やそのレベルの数は予め適宜に定めておけばよいが、一つの強度値レベルには後述する度数分布を判定するのに十分な数の行数が含まれるようにすることが望ましい。ピークマトリクスの全ての行がいずれかの強度値レベルに分類されると、レベル識別部は、複数段階の強度値レベル毎に、その強度値レベルに含まれる複数の行中のピーク強度値に基づいて、その強度値レベルがグループ間の差異解析に有用であるか否か、つまりは差異がある又は差異がないと判定するのに信頼に足るか否かを調べる。

40

【0021】

具体的には、ピークマトリクスの行毎に、その行に含まれるピーク強度値に基づいてグループ間に有意差があるか否かを判断するための指標値、例えば統計的仮説検定におけるp値を算出する。そして、一つの強度値レベルに分類された複数の行に対応する上記指標値の度数分布に基づいて、その強度値レベルが差異解析に有用であるか否かを判定し、最終的に複数段階の強度値レベルの中で差異解析に有用でない解析不適レベルを識別する。当該強度値レベル内に有意差のあるピークが存在しない場合には、上記p値の度数分布は

50

一樣になる。したがって、p値がほぼ一樣に分布している場合、有意差を示せるピークが含まれない可能性が高いと判断することができ、解析不適レベルであるとする事ができる。

【0022】

こうした処理を実施するために、本発明に係る質量分析データ処理装置において前記レベル識別部は、

c1)前記ピークマトリクスの行毎に統計的仮説検定におけるp値を計算するp値計算部と、

c2)強度値レベル毎に、一つの強度値レベルに分類された複数の行に対応するp値の値に関する度数分布を求める度数分布取得部と、

c3)強度値レベル毎に、p値の値に関する度数分布からその度数の一樣性を判定し、度数の一樣性が高い強度値レベルを解析不適レベルであると判断する判定部と、

を含む構成とすることができる。

【0023】

解析不適レベルに含まれる行中のデータ(ピーク強度)は差異解析に有用でないから、マトリクス修正部は、元のピークマトリクスから解析不適レベルに属する行を削除したフィルタリング済みピークマトリクスを取得する。そして、このフィルタリング済みピークマトリクスに対する統計的仮説検定を行うことで、グループ間の差異解析を行うようにすればよい。統計的仮説検定としては例えば、グループの数が二である場合にはt検定やU検定、グループの数が三以上である場合にはANOVAを用いることができる。

【0024】

通常、解析不適レベルであると判定される強度値レベルは、比較的信号強度が高いためピーク検出の際には排除できなかったノイズピークに対応する行やベースラインノイズが重畳しているピークに対応する行などを含んでいることが多い。こうした行はピーク検出の際の基準によってピークマトリクスに入る場合もあれば入らない場合もあり、それによってピークマトリクスの行数は変わることになる。一方、そうした行がピークマトリクス中に存在すれば、そうした行は解析不適レベルに含まれる行として除去される可能性が高い。したがって、本発明に係る質量分析データ処理装置によれば、与えられるマススペクトルデータが同じである限り、元のピークマトリクスの行数に拘わらず、つまりはピーク検出の際の基準に拘わらず、差異解析に有用であるデータのみを含む、ほぼ決まった行数のピークマトリクス(フィルタリング済みピークマトリクス)を差異解析に供することができる。その結果、ピーク検出の際の基準がBH法等による検定結果に与える影響を低減することができる。

【0025】

なお、本発明に係る質量分析データ処理装置では、前記フィルタリング済みピークマトリクスに基づいてグループ間の差異解析を実行し、該差異解析において有意差がある行に対応するピーク又は質量電荷比を抽出する差異解析部、をさらに備える構成とすることもできる。

【0026】

また、本発明に係る質量分析装置は、上記本発明に係る質量分析データ処理装置を含む質量分析装置であって、

サンプルに対し質量分析を行うことでマススペクトルデータを取得する質量分析部と、複数のグループのいずれかに属する複数のサンプルに対し前記質量分析部で質量分析を実行することで得られたマススペクトルデータを格納するデータ記憶部と、

をさらに備え、前記データ記憶部に格納されているマススペクトルデータを前記質量分析データ処理装置のピークマトリクス作成部に与えることを特徴としている。

【0027】

ここで、例えば質量分析部は、予め用意された多数のサンプルを自動的に交換しながら質量分析を順次実行するものとする事ができる。本発明に係る質量分析装置によれば、分析者が予め複数のサンプルを用意しさえすれば、その複数のサンプルに対する質量分析

10

20

30

40

50

、及び、質量分析により得られたデータに基づく差異解析を連続的に実行し、差異解析結果、つまりはグループ間で有意な差を示すマーカーの情報を得るようにすることができる。それにより、精度の高い差異解析を効率良く実行することができる。

【発明の効果】

【0028】

本発明に係る質量分析データ処理装置、質量分析データ処理方法、及び質量分析データ処理用プログラムによれば、分析者等により恣意的に決められるピーク検出のためのノイズレベルに左右されることなく、差異解析に意味のある信頼性の高いデータを差異解析に供することができる。それにより、元のマススペクトルデータが同じであれば、常に高い精度の差異解析を安定的に行うことができる。その結果として、複数のサンプルグループ間の差異の識別に有用なマーカー候補を的確に見つけることが可能となる。

10

【図面の簡単な説明】

【0029】

【図1】本発明に係る質量分析データ処理装置を用いた質量分析装置の一実施例の概略構成図。

【図2】本実施例の質量分析装置における差異解析の処理手順を示すフローチャート。

【図3】図2中の差異解析不適レベル判定処理（ステップS5）の詳細な処理手順を示すフローチャート。

【図4】ピークマトリクス概念図（a）及びピークマトリクスの一例を示す図（b）。

【図5】マトリクス行毎の代表ピーク強度値（Peak Intensity）とp値（p-value）との関係の一例を示すグラフ。

20

【図6】図5に示したグラフ中のデータに基づく強度値レベル毎のp値ヒストグラムを示す図。

【図7】図6に示したp値ヒストグラムに基づく強度値レベル毎の χ^2 検定結果を示す図。

【図8】差異解析不適レベル判定結果を用いたピークマトリクスの行フィルタリングの説明図であり、（a）はフィルタリング前のピークマトリクス、（b）はフィルタリング後のピークマトリクス。

【発明を実施するための形態】

【0030】

以下、本発明に係る質量分析データ処理装置を用いた質量分析装置の一実施例について、添付図面を参照して説明する。

30

【0031】

図1は本実施例の質量分析装置の概略構成図である。この質量分析装置は、例えば特定の疾病に罹患している多数の患者から採取した尿等の生体サンプルと多数の健常者から採取した尿等の生体サンプルとをそれぞれ質量分析することで得られたデータに基づいて、その特定の疾患に罹患していることを診断するのに有力なバイオマーカー候補を見つけるためのものである。

【0032】

本実施例の質量分析装置は、サンプルに対する質量分析を実行して所定の質量電荷比範囲に亘るイオン強度データつまりはマススペクトルデータを取得する質量分析装置本体1と、該質量分析装置本体1で収集されたデータを処理して差異解析を実施するデータ処理部2と、ユーザー（分析者）が何らかの入力や指示を行ったりするための入力部3と、解析結果等が表示される表示部4と、を備える。

40

【0033】

質量分析装置本体1はその構成を問わないが、例えば高い質量分解能、高い検出感度で以てマススペクトルデータが収集可能であるMALDI-TOFMSである。また、好ましくは、予め用意された多数のサンプルを自動的に交換しながら、各サンプルに対する質量分析を順次実行可能な装置とするとよい。予め多数のサンプルが用意されたサンプルプレートを二軸方向に移動させつつ、そのサンプルプレート上の多数のサンプルに対して順

50

番にレーザ光を照射して各サンプルの質量分析を行う、またさらには複数のサンプルプレートを自動的に交換して各サンプルの質量分析を行う構成のMALDI-TOFMSは、質量分析装置本体1として好適である。

【0034】

データ処理部2は、後述する特徴的なデータ処理を遂行するために、データ格納部21、ピーク検出部22、ピークマトリクス作成部23、マトリクス行レベル分け部24、差異解析不適レベル判定部25、マトリクス行フィルタリング部26、及び、差異解析部27、を機能ブロックとして備える。また、差異解析不適レベル判定部25は、さらに詳細な機能ブロックとして、p値計算部251、p値ヒストグラム作成部252、及びp値度数一様性判定部253、を含む。

10

【0035】

なお、一般に、データ処理部2の実体はパーソナルコンピュータ又はより高性能なコンピュータであり、該コンピュータに予めインストールされた専用のデータ処理ソフトウェアをコンピュータ上で実行することにより、上記のような各機能ブロックが具現化される。その場合、入力部3はコンピュータのキーボードやマウス等のポインティングデバイスであり、表示部4はモニタである。こうした構成では、コンピュータにインストールされたデータ処理ソフトウェアの全て又は一部が本発明に係る質量分析データ処理用プログラムに相当する。

【0036】

本実施例の質量分析装置では、質量分析装置本体1において所定のサンプルに対し質量分析が実行されることで得られた所定の質量電荷比範囲のマスペクトルデータは逐次データ処理部2に送られ、サンプル名等のサンプル識別情報、グループ名等のグループ識別情報が付与されてそれぞれデータ格納部21に格納される。ただし、図1に示している或る特定の質量分析装置本体1により得られたマスペクトルのみならず、他の質量分析装置において同様に得られたマスペクトルデータもデータ格納部21に格納することが可能である。即ち、データ処理部2において処理対象である多数のサンプルに対するマスペクトルデータは、1台の質量分析装置で得られたデータでもよいし、異なる複数の質量分析装置で得られたデータでもよい。いずれにしても、ここでは、二つのグループのいずれかに属する多数のサンプルに対するマスペクトルデータがデータ格納部21に格納される。

20

30

【0037】

例えば分析者が入力部3から処理対象のデータを指定したうえで処理の実行を指示すると、これを受けて、ピーク検出部22は指定されたデータをデータ格納部21から順次読み出すことで取得する(ステップS1)。そして、個々のマスペクトルデータに対し所定の基準に従ってマスペクトル上のピークを検出し、検出された各ピークの位置(質量電荷比m/z値)とピーク強度値とを求める。ピーク検出のアルゴリズムは従来から知られている適宜の手法を用いることができ、例えばピーク状波形の信号強度が予め定めておいた閾値を超えるものをピークとして検出すればよい。そして、マスペクトル毎つまりはサンプル毎に、検出されたピークの質量電荷比値とピーク強度値との組を集めたピークリストを作成する(ステップS2)。

40

【0038】

上述したように元のマスペクトルデータにはサンプル識別情報とグループ識別情報が付与されているから、個々のサンプルリストにもサンプル識別情報とグループ識別情報が付与される。ここでは、二つのグループを識別するグループ識別情報を「Group1」と「Group2」とし、各サンプルを識別するサンプル識別情報を「Sample1-1」、「Sample1-2」、...、「Sample2-1」、「Sample2-2」、...、とする。

【0039】

ピークマトリクス作成部23は、ステップS2において作成された全てのピークリストを整理してピークマトリクスを作成する(ステップS3)。具体的には、図4(a)に示したように、全てのピークリストに挙げられているピークの質量電荷比値を縦方向(列方

50

向)に並べる一方、サンプルを横方向(行方向)に並べ、ピーク強度値を要素としてピークマトリクスを作成する。或るサンプルにおいてピークが存在し他のサンプルではピークが存在しない質量電荷比値の行において、その「他のピーク」に該当する要素は存在しないから、その要素であるピーク強度はゼロとすればよい。図4(b)はこうして作成されるピークマトリクスの一例である。

【0040】

続いてマトリクス行レベル分け部24は、作成されたピークマトリクスの各行を、そのマトリクス行に含まれる少なくとも一つのピーク強度値に基づいて複数段階の強度値レベルのいずれかに分類し、各マトリクス行に強度値レベルを示す情報(例えばフラグ)を付加する(ステップS4)。例えば一つの方法として、一つのマトリクス行に含まれる複数のピーク強度値の中で最大のものを探索し、それをそのマトリクス行の代表ピーク強度値として該代表ピーク強度値に基づいて複数の強度値レベルのいずれかへ分類する。或いは、一つのマトリクス行に含まれる全てのピーク強度値の平均値や中央値を代表ピーク強度値としてもよい。

【0041】

図5は、マトリクス行中の最大ピーク強度を代表ピーク強度値としたときの、一つのピークマトリクスにおける代表ピーク強度値(横軸のPeak Intensity)の分布を示すグラフである。ただし、このグラフの縦軸のp値(p-value)についてはあとで説明する。この例では、代表ピーク強度値は $1E+03 \sim 1E+08$ の範囲に分布している。そこで、ここでは図中に示したように、 $1E+03 \sim 1E+04$ 、 $1E+04 \sim 1E+05$ 、 $1E+05 \sim 1E+06$ 、 $1E+06 \sim 1E+07$ 、 $1E+07 \sim 1E+08$ 、という5段階の強度値範囲であるレベル(Level 1~Level 5)を定めて、各マトリクス行の代表ピーク強度値がいずれの強度値レベルに入るのかを判定することで各マトリクス行をレベル分けするようにしている。そして、図8(a)に示すように各マトリクス行に強度値レベルを示す情報を付加する。

【0042】

そのあと差異解析不適レベル判定部25は、上記複数段階の強度値レベル毎に、その強度値レベルに含まれるマトリクス行中のピーク強度値に基づいて、当該強度値レベルが差異解析のために有用な強度値レベルであるか否かを判定し、差異解析に有用でない、即ち、差異解析を実施したときに意味のある結果が得られない差異解析不適レベルを見つける(ステップS5)。

【0043】

ここで、このステップS5における差異解析不適レベル判定処理を、図3に示すフローチャートを参照して詳述する。

まず、p値計算部251は、レベル分け済みの各マトリクス行について、そのマトリクス行に含まれる全てのピーク強度値に基づき、グループ間の差異についての統計的仮説検定を行い有意差の指標値であるp値を計算する(ステップS51)。統計的仮説検定としては後述するステップS7における差異解析と同じ手法を用いればよく、例えばt検定を利用することができる。そして、全てのマトリクス行についてそれぞれp値を求める。

【0044】

先に説明した図5に示したグラフの縦軸はこのとき計算されるp値を示したものであり、図5のグラフ中の一つのプロットは、ピークマトリクス中の或る一つのマトリクス行を示している。例えば図5において点線で囲んだ矩形の領域中に存在する複数のプロットは、強度値の範囲が $1E+03 \sim 1E+04$ である強度値レベルLevel 1に含まれるマトリクス行を示している。

【0045】

次にp値ヒストグラム作成部252は、5段階の強度値レベル毎に、ステップS51で算出されたp値を複数の値の範囲に分けてその範囲毎に度数(頻度)を求めることでp値の値の度数分布を示すp値ヒストグラムを作成する(ステップS52)。図6は図5に示したデータに基づいて作成されたp値ヒストグラムであり、横軸がp値の値の範囲、縦軸が度数である。図5ではp値が小さい(例えば0.1以下)領域にプロットが比較的集中

10

20

30

40

50

していることが感覚的に分かるが、図6に示したp値ヒストグラムでは度数の高さからそのことが確認できる。

【0046】

p値度数一様性判定部253は、強度値レベル毎に、p値ヒストグラムにおける度数分布の一様性を判定し、それに基づいて、差異解析に適さない強度値レベル、つまりは差異解析不適レベルを見つける(ステップS53)。

本例では、具体的に以下のような処理を行う。図7は、図6に示したp値ヒストグラムにおける度数を数値で示したものである。ここでは度数分布が一様であるか非一様(片寄っている)であるかを判定するために、統計的仮説検定の一つである χ^2 (カイ二乗)検定を用いる。即ち、度数分布が一様であるという帰無仮説を検定するために χ^2 検定を用いてそのp値を求めると、図7中の最下行に示すようになる。ここでは、Level1のp値のみが0.05よりも大きいことから、Level1における度数分布は一様であるという帰無仮説は棄却されない。一方、Level2以上では度数分布は一様でない、つまりは何らかの偏りがみられると判定される。

【0047】

マトリクス行におけるp値はグループ間で有意差があるか否かを示す指標値であるから、p値の度数分布の一様性が高いということはグループ間の有意差の判断のばらつきが大きいこと、即ち、差異解析における判断の信頼性が低いことを意味している。そこで、p値の度数分布の一様性が高いと判定された強度値レベルについては差異解析不適レベルとする。図7に示した例では、Level1のみがp値の度数分布の一様性が高いと判定され、この強度値レベルのみが差異解析不適レベルとされる。

【0048】

差異解析不適レベルに含まれるデータ(ピーク強度値)は差異解析において信頼性が低いと推定される。そこで、マトリクス行フィルタリング部26は、ステップS4において強度値レベルを示す情報が付加されたピークマトリクスから差異解析不適レベルに属するマトリクス行を削除し、行フィルタリング済みピークマトリクスを取得する(ステップS6)。図8(b)は図8(a)に示したピークマトリクスから差異解析不適レベルであるLevel1に属するマトリクス行を削除した場合の例である。

【0049】

当然、差異解析不適レベルが一つもない場合もあれば、複数存在する場合もあり得る。また、仮に或る差異解析不適レベルよりも強度値が低い別の強度値レベルが差異解析不適レベルと判定されていない場合であっても、強度値がその差異解析不適レベル以下である強度値レベルは全て差異解析不適レベルとするとよい。例えば、上記例においてLevel1が差異解析不適レベルと判定されずにLevel2が差異解析不適レベルと判定された場合には、Level1、Level2共に差異解析不適レベルとするとよい。これは、強度値が小さいにも拘わらず差異解析不適レベルと判定されなかった強度値レベルは偶然、差異解析不適レベルと判定されなかった可能性が高いと推測されるためである。

【0050】

差異解析部27は行フィルタリング済みピークマトリクスを受け取り、該ピークマトリクスに対して例えばt検定やU検定などの統計的仮説検定による多重比較検定を実行することでグループ間の差異解析を行う。この差異解析の方法自体は従来と同じであり、例えば上記BH法によりFDRを所定の有意水準以下に抑える条件の下で差異解析を行えばよい。そして、その差異解析によって求めた、グループ間で有意差があるマトリクス行つまりピークを特定し、例えばそのピークの質量電荷比をバイオマーカー候補として表示部4を通し分析者に提示する(ステップS7)。

【0051】

上述したように、分析者が定めたノイズレベルの閾値を基準として信号強度を判定してピーク検出を行う場合、元のマススペクトルデータは同じであってもその閾値によってピークマトリクスの行数が変化する。これに対し、差異解析不適レベルの判定は元のデータ(ピーク強度値)に基づくものであり基本的には分析者等による恣意的な操作の影響を受

10

20

30

40

50

けないため、通常、元のマススペクトルデータが同じであれば差異解析をやり直しても行フィルタリング済みピークマトリクスの行数は常に一定である。したがって、分析者等の恣意的な操作による差異解析結果の変動が生じにくい。また、ステップS5～S6の処理により差異解析における信頼性が低いと推定されるデータが除去されるため、差異解析の正確性が従来よりも向上する。

【0052】

以上のようにして、本実施例の質量分析システムでは、それぞれ多数のサンプルが属している二つのグループの差異に寄与するマーカーを精度良く探索することができる。

また、差異解析を行いたいグループの数が三以上である場合には、統計的仮説検定としてt検定やu検定の代わりに多群検定に適したANOVAを用いればよい。ANOVAでもt検定等と同様にp値を求めることができるから、上述した手順によりピークマトリクスから三以上のグループの差異解析に適さない行を削除したうえで差異解析を実行することができる。

10

【0053】

なお、上記処理の中で強度値レベルの数や幅（数値の範囲）は適宜に定めることが可能である。ただし、差異解析不適レベルを判定する際に一つの強度値レベルの中に含まれるマトリクス行が少なすぎると正確な判定が困難になるから、強度値レベルの数を多くしすぎたりその幅を狭くしすぎたりしないように注意を要する。一般的には、一つの強度値レベルに少なくとも10程度以上のマトリクス行数が入るように、強度値レベルの数や幅が設定されていることが望ましい。

20

【0054】

また、図3に示したフローチャートによる差異解析不適レベルの判定手法は一例であり、マトリクス行毎にグループ間の有意差を判断し得る又は有意差の判断に関連する指標値を求め、強度値レベル毎にその指標値の値の度数分布の一樣性や分散性を判断することで、差異解析に有意でない強度値レベルを見つけることができる。

【0055】

また、上でも述べたように、ピークマトリクスの行と列とは入れ替える、つまりは縦方向（列方向）にサンプル情報、横方向（行方向）に質量電荷比値をとるようにすることができることは明白である。

【0056】

また、上記実施例はあくまでも本発明の一例にすぎず、本発明の趣旨の範囲で適宜変形、修正、追加等を行っても本願特許請求の範囲に包含されることは当然である。例えば、上記実施例において、解析対象であるマススペクトルデータはMRM（Multiple Reaction Monitoring）測定における定量用のMSⁿスペクトルデータであっても構わない。

30

【符号の説明】

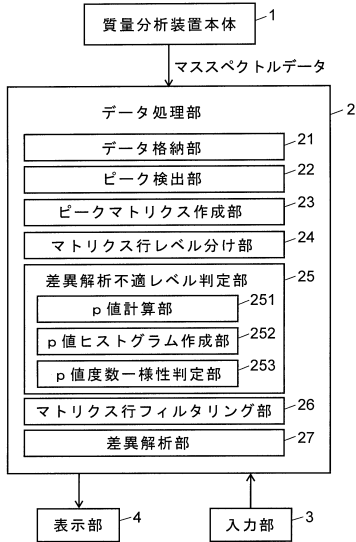
【0057】

- 1 ... 質量分析装置本体
- 2 ... データ処理部
- 2 1 ... データ格納部
- 2 2 ... ピーク検出部
- 2 3 ... ピークマトリクス作成部
- 2 4 ... マトリクス行レベル分け部
- 2 5 ... 差異解析不適レベル判定部
- 2 5 1 ... p値計算部
- 2 5 2 ... p値ヒストグラム作成部
- 2 5 3 ... p値度数一樣性判定部
- 2 6 ... マトリクス行フィルタリング部
- 2 7 ... 差異解析部
- 3 ... 入力部
- 4 ... 表示部

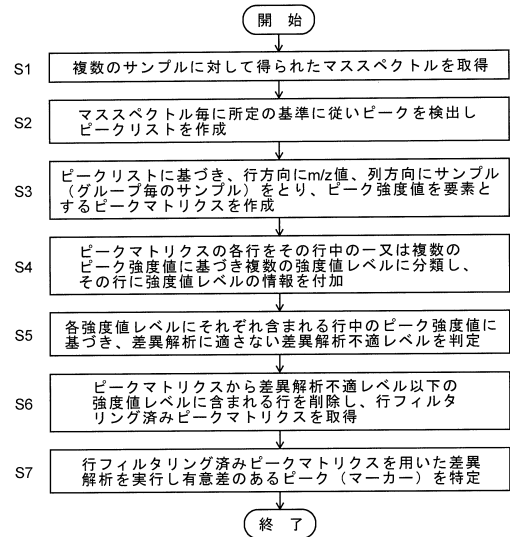
40

50

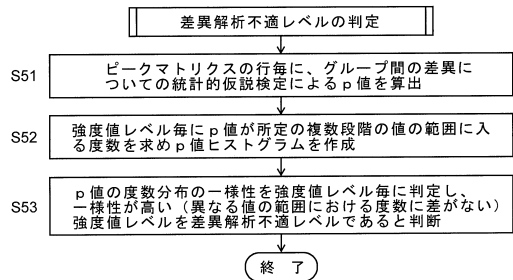
【図1】



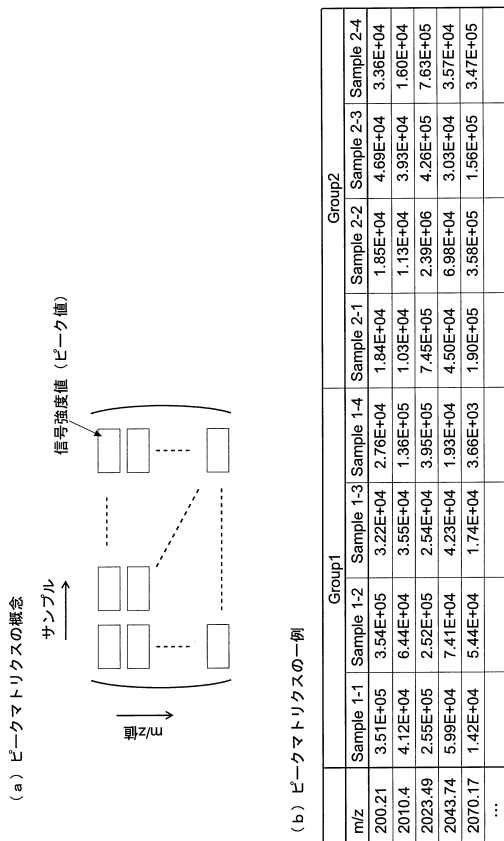
【図2】



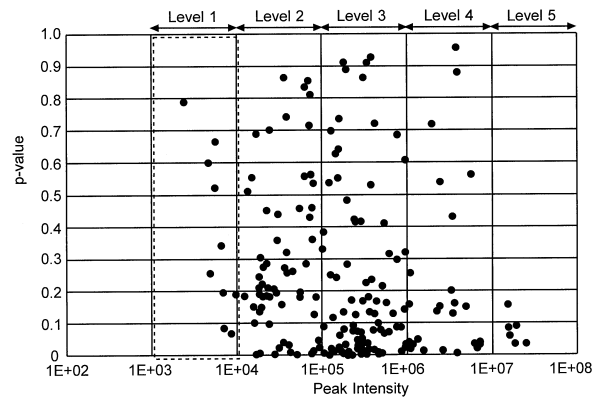
【図3】



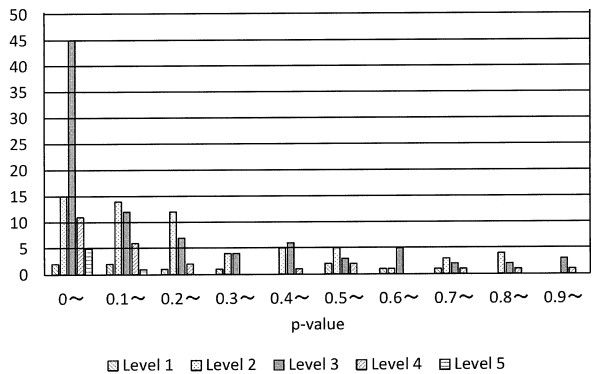
【図4】



【図5】



【図6】

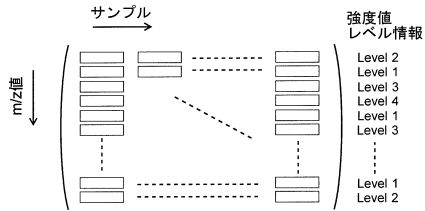


【 図 7 】

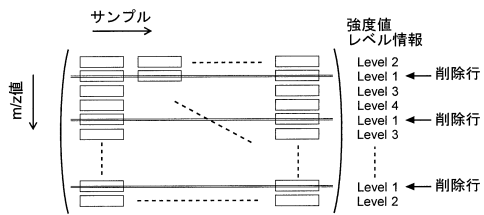
p 値範囲	Level 1	Level 2	Level 3	Level 4	Level 5
0~	2	15	45	11	5
0.1~	2	14	12	6	1
0.2~	1	12	7	2	0
0.3~	1	4	4	0	0
0.4~	0	5	6	1	0
0.5~	2	5	3	2	0
0.6~	1	1	5	0	0
0.7~	1	3	2	1	0
0.8~	0	4	2	1	0
0.9~	0	0	3	1	0
χ^2 test	0.739918	4.44E-06	2.62E-32	2.55E-06	2.29E-05

【 図 8 】

(a) レベル分け済みピークマトリクス



(b) 行フィルタリング済みピークマトリクス



フロントページの続き

- (56)参考文献 特開2014-202582(JP,A)
特表2007-533997(JP,A)
特表2008-547005(JP,A)
国際公開第2008/143494(WO,A1)

- (58)調査した分野(Int.Cl., DB名)
G01N 27/60-70、92