(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4298101号
(P4298101)
(45)発行日　平成21年7月15日(2009.7.15)　(24)登録日　平成21年4月24日(2009.4.24)

(51)Int.Cl.　　F I
G06Q 10/00 (2006.01)　　G06F 19/00 100
C12N 15/09 (2006.01)　　C12N 15/00 A
G06F 17/30 (2006.01)　　G06F 17/30 170F

請求項の数 18　(全 21 頁)

(21)出願番号　特願平11-371434
(22)出願日　平成11年12月27日(1999.12.27)
(65)公開番号　特開2001-178463(P2001-178463A)
(43)公開日　平成13年7月3日(2001.7.3)
審査請求日　平成14年6月7日(2002.6.7)
審判番号　不服2004-17732(P2004-17732/J1)
審判請求日　平成16年8月26日(2004.8.26)

(73)特許権者　000233055
日立ソフトウエアエンジニアリング株式会社
東京都品川区東品川四丁目１２番７号
(74)代理人　100091096
弁理士　平木　祐輔
(74)代理人　100102576
弁理士　渡辺　敏章
(72)発明者　渡辺　恒彦
神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

最終頁に続く

(54)【発明の名称】類似発現パターン抽出方法及び関連生体高分子抽出方法

(57)【特許請求の範囲】
【請求項１】
発現パターンデータを記憶し表示部を備えた処理装置を用い、生体高分子に対する複数の実験ケースと発現量との関係を表す検索対象発現パターンの集合の中から検索発現パターンに類似するものを抽出する類似発現パターン抽出方法において、
前記検索対象発現パターン及び前記検索発現パターンに対して実験ケース間での実験環境の違いによる発現データのバラつきを補正する正規化を行う工程と、
発現量を縦軸にとり実験ケースを横軸にとって、前記検索対象発現パターンの集合を前記表示部の検索対象発現パターン表示領域に重ねて表示すると共に、前記検索発現パターンを前記表示部の検索発現パターン表示領域に表示する工程と、
前記検索発現パターンの横軸方向の一部分の区間を選択する工程と、
前記選択された横軸方向の区間において実験ケース毎に所定量の発現量を前記検索発現パターンに加減することによって前記縦軸方向に一定の幅を持った帯状のパターンを発生する工程と、
前記検索対象発現パターン表示領域内に検索区間を設定する工程と、
前記帯状のパターンを前記検索対象発現パターン表示領域内に設定された前記検索区間内で縦軸及び／又は横軸方向に平行移動しながら、前記帯状のパターンの中にユーザが指定した一定割合以上入る検索対象発現パターンを抽出する工程と、
前記抽出された検索対象発現パターンの部分と前記検索発現パターンの前記選択された横軸方向の区間との類似度を予め選択された計算方法に従って計算する工程と、

抽出された前記検索区間における前記検索対象発現パターンと前記選択された区間における前記検索発現パターンを前記表示部にオーバーラップして表示すると共に、抽出した検索対象発現パターンの詳細表示として、パターン間の類似度、遺伝子名、及び遺伝子に関する情報を前記計算された類似度の高い遺伝子の順に並び替えて表示する工程と

を有することを特徴とする類似発現パターン抽出方法。

【請求項２】

請求項１記載の類似発現パターン抽出方法において、前記検索区間内で、前記帯状のパターンを前記縦軸方向又は横軸方向に平行移動しながら前記帯状のパターンの中にユーザが指定した一定割合以上入る検索対象発現パターンを抽出する工程を、前記帯状のパターンを前記横軸方向又は縦軸方向に所定距離だけずらしながら反復することを特徴とする類似発現パターン抽出方法。

【請求項３】

請求項１又は２記載の類似発現パターン抽出方法において、前記生体高分子は遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質であることを特徴とする類似発現パターン抽出方法。

【請求項４】

請求項１～３のいずれか１項記載の類似発現パターン抽出方法において、前記実験ケースは時系列に沿った実験であることを特徴とする類似発現パターン抽出方法。

【請求項５】

請求項１～４のいずれか１項記載の類似発現パターン抽出方法において、前記実験ケースは個体の種別であることを特徴とする類似発現パターン抽出方法。

【請求項６】

請求項１～４のいずれか１項記載の類似発現パターン抽出方法において、前記実験ケースは個体の部位であることを特徴とする類似発現パターン抽出方法。

【請求項７】

請求項１～４のいずれか１項記載の類似発現パターン抽出方法において、前記実験ケースは人工的な条件の印加と無印加であることを特徴とする類似発現パターン抽出方法。

【請求項８】

請求項１～４のいずれか１項記載の類似発現パターン抽出方法において、前記実験ケースは時系列に沿った実験、異なる種、個体の部位、及び人工的な条件の印加と無印加のうちの複数を複合したものであることを特徴とする類似発現パターン抽出方法。

【請求項９】

請求項１記載の類似発現パターン抽出方法において、検索対象発現パターンの集合の中から複数の検索発現パターンに類似するものをそれぞれ抽出することを特徴とする類似発現パターン抽出方法。

【請求項１０】

発現パターンデータを記憶し表示部を備えた処理装置を用い、生体高分子に対する複数の実験ケースと発現量との関係を表す発現パターンどうしを比較して、検索対象生体高分子の中から検索生体高分子と発現において関連を有する生体高分子を抽出する関連生体高分子抽出方法において、

前記検索対象生体高分子及び前記検索生体高分子の発現パターンに対して実験ケース間での実験環境の違いによる発現データのバラつきを補正する正規化を行う工程と、

発現量を縦軸にとり実験ケースを横軸にとって、前記検索対象生体高分子の発現パターンの集合を前記表示部の検索対象生体高分子発現パターン表示領域に重ねて表示すると共に、前記検索生体高分子の発現パターンを前記表示部の検索生体高分子発現パターン表示領域に表示する工程と、

前記検索生体高分子の発現パターンの横軸方向の一部分の区間を選択する工程と、

前記選択された横軸方向の区間において実験ケース毎に所定量の発現量を前記検索生体高分子の発現パターンに加減することによって前記縦軸方向に一定の幅を持った帯状のパターンを発生する工程と、

前記検索対象生体高分子発現パターン表示領域内に検索区間を設定する工程と、
前記帯状のパターンを前記検索対象生体高分子発現パターン表示領域内に設定された前記検索区間内で縦軸及び／又は横軸方向に平行移動しながら、前記帯状のパターンの中にユーザが指定した一定割合以上入る検索対象発現パターンを有する検索対象生体高分子を抽出する工程と、
前記抽出された検索対象生体高分子の発現パターンの部分と前記検索生体高分子の発現パターンの前記選択された横軸方向の区間との類似度を予め選択された計算方法に従って計算する工程と、
抽出された前記検索区間における前記検索対象生体高分子の発現パターンと前記選択された区間における前記検索生体高分子の発現パターンを前記表示部にオーバーラップして表示すると共に、抽出した検索対象生体高分子の詳細表示として、発現パターン間の類似度、生体高分子名、及び生体高分子に関する情報を前記計算された類似度の高い生体高分子の順に並び替えて表示する工程と
を有することを特徴とする関連生体高分子抽出方法。
【請求項１１】
請求項１０記載の関連生体高分子抽出方法において、前記検索区間内で、前記帯状のパターンを前記縦軸方向又は横軸方向に平行移動しながら前記帯状のパターンの中にユーザが指定した一定割合以上入る検索対象発現パターンを有する検索対象生体高分子を抽出する工程を、前記帯状のパターンを前記横軸方向又は縦軸方向に所定距離だけずらしながら反復することを特徴とする関連生体高分子抽出方法。
【請求項１２】
請求項１０又は１１記載の関連生体高分子抽出方法において、前記生体高分子は遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質であることを特徴とする関連生体高分子抽出方法。
【請求項１３】
請求項１０～１２のいずれか１項記載の関連生体高分子抽出方法において、前記実験ケースは時系列に沿った実験であることを特徴とする関連生体高分子抽出方法。
【請求項１４】
請求項１０～１２のいずれか１項記載の関連生体高分子抽出方法において、前記実験ケースは個体の種別であることを特徴とする関連生体高分子抽出方法。
【請求項１５】
請求項１０～１２のいずれか１項記載の関連生体高分子抽出方法において、前記実験ケースは個体の部位であることを特徴とする関連生体高分子抽出方法。
【請求項１６】
請求項１０～１２のいずれか１項記載の関連生体高分子抽出方法において、前記実験ケースは人工的な条件の印加と無印加であることを特徴とする関連生体高分子抽出方法。
【請求項１７】
請求項１０～１２のいずれか１項記載の関連生体高分子抽出方法において、前記実験ケースは時系列に沿った実験、異なる種、個体の部位、及び人工的な条件の印加と無印加のうちの複数を複合したものであることを特徴とする関連生体高分子抽出方法。
【請求項１８】
請求項１０記載の関連生体高分子抽出方法において、検索対象生体高分子の集合の中から、複数の検索生体高分子の発現パターンを用いて、前記各検索生体高分子と発現において関連を有するものをそれぞれ抽出することを特徴とする関連生体高分子抽出方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、遺伝子等の生体高分子と発現に関して関連を有する生体高分子を大量の生体高分子群の中から抽出する方法に関する。
【０００２】

【従来の技術】
さまざまな種においてゲノム配列が決定されるのに伴い、進化に対応するとみられる遺伝子を見つけ出し、どの生物も共通に持っていると考えられる遺伝子の集合を探したり、それから逆に種毎に個別な特徴を推測するなど、種間の遺伝子の違いから何かを見出そうとする、いわゆるゲノム比較法が盛んに行われてきた。
【０００３】
しかし近年、ＤＮＡチップやＤＮＡマイクロアレイなどのインフラストラクチャの発達によって、分子生物学の興味は、種間の情報から種内の情報へ、すなわち同時発生解析へと移りつつあり、これまでの種間の比較と合わせて、情報の抽出から関連付けの場が大きく広がりを持ち始めている。
例えば、既知の遺伝子と同一の発現パターンを示す未知の遺伝子が見つかれば、その遺伝子には既知の遺伝子と同様の機能があると類推できる。これら遺伝子や蛋白質そのものの機能的な意味付けは、機能ユニットや機能グループといった形で研究されている。また、それらの間の相互作用も、既知の酵素反応データや物質代謝データとの対応付けによって、あるいはより直接的に、ある遺伝子を破壊あるいは過剰反応させ、その遺伝子の発現をなくすか、あるいは多量に発現させ、その遺伝子の直接的及び間接的影響を、全遺伝子の発現パターンを調べることによって解析している。
【０００４】
ここで、ある遺伝子の発現パターンとは、横軸に実験ケース、縦軸に発現量をとり、その遺伝子に対して一連の実験ケースで得られた発現量の値をつなげて曲線（あるいは折れ線）グラフとして表したものである。なお、遺伝子だけでなくＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質などの生体高分子についても同様の発現パターンを考えることができるが、ここでは遺伝子の発現パターンを例にとって説明する。横軸にとる実験ケースの具体例としては、時系列、生物の体の部位、生物種、塩基配列の一部、遺伝子などがある。
【０００５】
実験ケースにおいて時系列とした発現パターンを解析した例としては、例えばスタンフォード大学のP. Brownらのグループによるイースト菌の発現解析が挙げられる（Michel B. Eisen et. al.: Cluster analysis and display of genome-wide expression patterns: Proc. Natl. Acad. Sci. 1998 Dec 8, 95(25), 14863-8）。彼らは、ｃｄ６変異株と呼ばれる遺伝子を用い、遠心分離法を適用したときの発現過程、出芽時期における発現過程、高熱ショックを与えたときの発現過程、低熱ショックを与えたときの発現過程、ジーオーキーシーシフト法を適用したときの発現過程などの各実験における時系列の遺伝子の発現データを組み合わせ、発現パターン同士をクラスタリングすることにより、遺伝子の機能を特定するのに成功したことが知られている。
【０００６】
遺伝子の発現解析法のひとつとして、図２４（ａ）に示すような大量の遺伝子（検索対象遺伝子）群の発現パターンの中から、図２４（ｂ）に示すように任意に選択した遺伝子（検索遺伝子）の発現パターンによく類似したパターンを有する遺伝子を抽出する方法がある。抽出された遺伝子は、検索遺伝子の属する機能グループや機能ユニットの候補とすることができる。
【０００７】
この従来の方法は、検索の様子を図２４（ｃ）に模式的に示し、検索結果を図２４（ｄ）に示すように、検索遺伝子の発現パターンとそのパターン全体において類似した発現パターンをもつ遺伝子を抽出する方法である。具体的には一つの遺伝子の発現過程を一つのベクトル（実験ケースと同じ数だけ独立した軸を有する多次元空間において、各実験ケースにおける発現量をその実験ケースに相当する軸方向の成分として持つベクトル）に対応させ、遺伝子間の類似性をベクトル間の（非）類似度によって比べている。また、実際の遺伝子データではなくユーザが任意に曲線データ（発現パターンデータ）を用意し、これを基準として遺伝子を抽出する方法もある。 これらの縦軸の発現量は遺伝子の数量に関す

る増幅割合を表すものであり、具体的な測定値は実験方法に依存するが、例えばＤＮＡチップ上でハイブリダイズした遺伝子を標識している標識蛍光物質から発せられる蛍光の強度、標識化学発光物質から発せられる化学発光の強度、ＤＮＡチップ上に付着した遺伝子に発生する化学反応起因の電気的シグナルを電極により検出した値、ハイブリダイズした遺伝子をガス化しその飛行時間を計測することによって得た質量分析の値などを発現量の指標とすることができる。
などを発現量の指標とすることができる。

【０００８】

【発明が解決しようとする課題】

ところが、前述の従来の方法では、任意に選択した検索遺伝子の発現パターンと検索対象遺伝子の発現パターンとが、そのパターンの全ての部分（全ての実験ケース）において類似している遺伝子しか抽出できなかった。
例えば図２２（ａ）に示すように、検索対象遺伝子の発現パターンデータに実験環境の違いによる測定誤差が含まれていた場合、従来の方法ではこれを検索遺伝子と類似した発現パターンであると認識できなかった。また、例えば図２２（ｂ）に示すように、時系列のある区間（ある連続する実験ケースの区間）に限り複数の遺伝子が同じ機能のために協調して発現し、その他の区間では異なる機能のために別々に発現している場合、ある区間では同じ形状の曲線が集まっているように見える。このとき、従来の方法では、このようにある区間で同じ形状となっている曲線群を抽出できなかった。

【０００９】

また、遺伝子の発現調節現象は、ある遺伝子の発現が別の遺伝子の発現を誘導したり、あるいは発現を阻害するなど、一連のカスケードで構成されている。
ここでカスケードとは、図２３に模式的に示すように、遺伝子１は遺伝子２の発現を誘導し、さらに、遺伝子２は遺伝子３、遺伝子３は遺伝子４を誘導するというように複数の遺伝子が連鎖的に発現する現象のことである。そしてさらに、これらカスケードが複合することにより複雑なネットワークが形成される。このような遺伝子のカスケードでは、時間の経過と共に複数の遺伝子発現のピークが連なってみられ、それらの発現パターンは非常に類似した形状をもつことがある。図２２（ｄ）も、このカスケードの一部分を表した図である。従来の方法では、このような遺伝子のカスケードを検出することができなかった。

【００１０】

これに加え、図２２（ｅ）に示すような、遺伝子の発現を抑制するように働く遺伝子発現パターンや、図２２（ｃ）に示すように協調して働くが発現量の度合いに一定量の大きな差がみられる遺伝子発現パターン、また、図２２（ｆ）に示すように常に一定の倍率をもって発現がみられる遺伝子発現パターン等を検出することができなかった。

【００１１】

本発明は、このような従来技術の問題点に鑑み、遺伝子等の発現データ特有の性質を考慮した類似発現パターン抽出方法及び関連生体高分子抽出方法を提供することを目的とする。特に、検索対象となる発現パターンが測定誤差を含む場合の解析、時系列のある区間に限り複数の生体高分子（例えば遺伝子）が同じ機能のために協調して発現し、その他の区間では異なる機能のために別々に発現している場合の解析、あるいは遺伝子のカスケードの解析などを支援することのできる類似発現パターン抽出方法及び関連生体高分子抽出方法を提供することを目的とする。

【００１２】

【課題を解決するための手段】

本発明ではまず、検索する遺伝子等の発現パターンのうち利用者が興味ある発現を示す実験ケース区間を図２（ａ），（ｂ）のように検索の範囲として部分的に選択する。そして、図２（ｃ）に示すように、この選択された発現パターンの一部分を検索対象遺伝子等の発現パターンと比べることにより、少なくとも一部分が類似した発現パターンを持つ遺伝子等を検索対象遺伝子群等の中から検索する。この方法を適用することで、例えば図２２

（ａ）の場合には誤差の範囲を検索対象外とすることで類似遺伝子等を抽出することができ、図２２（ｂ）の場合においても、同様な発現過程を示す範囲のみを検索対象とすることで、同じ発現過程を示す部分を抽出することができる。

【００１３】

また、図２（ｄ）に示すように、この部分的に取り出した曲線（発現パターンの一部）を任意に縦軸（発現量軸）方向又は横軸（実験ケース軸）方向に平行移動し、移動後の曲線（発現パターンの一部）と検索対象遺伝子等の発現パターン曲線とを重ねあわせることにより、図２（ｅ）に示すように、類似したパターン形状を有する発現パターンを検索できる。さらに、部分的に取り出した曲線（発現パターン）を任意に加工し、例えば抑制遺伝子ならば曲線を上下反転する、常に一定の倍率をもって発現がみられる遺伝子ならば曲線を縦軸方向に拡大・縮小するなどし、この加工した曲線（発現パターン）を用いることにより、より広い意味での類似パターンを持つ遺伝子曲線を検索できる。

【００１４】

検索遺伝子等の発現パターンの選択部分とこれに対応する検索対象遺伝子等の発現パターンのパターン形状を比較する手段としては、検索遺伝子等の発現パターン曲線の選択部分の発現量の値に任意の幅をもたせた帯を作り、検索対象遺伝子群等の発現パターンからこの帯に含まれる発現パターンを有する遺伝子等を取り出す方法、検索遺伝子等と検索対象遺伝子等との類似度を算出する方法、また、これらを組み合わせ、帯に含まれる遺伝子等の類似度を算出する方法を適用することができる。

【００１５】

すなわち、本発明による類似発現パターン抽出方法は、生体高分子に対する複数の実験ケースと発現量との関係を表す検索対象発現パターンの集合の中から検索発現パターンに類似するものを抽出する類似発現パターン抽出方法において、検索発現パターンの一部分を取り出し、取り出したパターン部分のパターン形状と類似するパターン形状を含む検索対象発現パターンを抽出することを特徴とする。

【００１６】

取り出したパターン部分に加工処理を施し、加工されたパターン部分のパターン形状と類似するパターン形状を含む検索対象発現パターンを抽出することもできる。加工処理としては、取り出したパターン部分を縦軸（発現量軸）に対して上下反転する処理、あるいは取り出したパターン部分の縦軸（発現量軸）方向あるいは横軸（実験ケース軸）方向の倍率を変化させる処理などが有効である。

【００１７】

類似するパターン形状の判定に当たっては、検索発現パターンの一部分を取り出したパターン部分あるいはそれを加工したパターン部分に対し実験ケース毎に所定量の発現量を加減（加算・減算）することによって縦軸（発現量軸）方向に一定の幅を持った帯状のパターンを発生し、パターンの一部が当該帯状のパターンの中に全て入る検索対象発現パターンを抽出するようにする。

また、帯状のパターンを縦軸方向あるいは横軸方向に平行移動しながら、パターン部分が当該帯状のパターンの中に全て入る検索対象発現パターンを抽出する。

この類似発現パターン抽出方法を適用する生体高分子は遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質とすることができる。また、実験ケースは時系列に沿った実験、個体の種別、個体の部位、人工的な条件の印加と無印加、あるいはこれらのうちの複数を複合したものとすることができる。

【００１８】

本発明による関連生体高分子抽出方法は、生体高分子に対する複数の実験ケースと発現量との関係を表す発現パターンどうしを比較して、検索対象生体高分子の中から検索生体高分子と発現において関連を有する生体高分子を抽出する関連生体高分子抽出方法において、検索生体高分子の発現パターンの一部分を取り出し、取り出したパターン部分のパターン形状と類似するパターン形状を含む発現パターンを有する検索対象生体高分子を抽出することを特徴とする。

【００１９】
取り出したパターン部分に加工処理を施し、加工されたパターン部分のパターン形状と類似するパターン形状を含む発現パターンを有する検索対象生体高分子抽出することもできる。加工処理としては、パターン部分を縦軸（発現量軸）方向に対して上下反転する処理、あるいはパターン部分の縦軸（発現量軸）方向あるいは横軸（実験ケース軸）方向の倍率を変化させる処理などが有効である。
【００２０】
類似するパターン形状の判定に当たっては、検索生体高分子の発現パターンの一部分を取り出したパターン部分あるいはそれを加工したパターン部分に対し実験ケース毎に所定量の発現量を加減することによって縦軸（発現量軸）方向に一定の幅を持った帯状のパターンを発生し、パターンの一部が当該帯状のパターンの中に全て入る発現パターンを有する検索対象生体高分子を抽出する。
また、帯状のパターンを縦軸方向あるいは横軸方向に平行移動しながら、パターンの一部が当該帯状のパターンの中に全て入る検索対象発現パターンを有する検索対象生体高分子を抽出する。
この関連生体高分子抽出方法を適用する生体高分子は遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質とすることができる。また、実験ケースは時系列に沿った実験、個体の種別、個体の部位、人工的な条件の印加と無印加、あるいはこれらのうちの複数を複合したものとすることができる。
【００２１】
【発明の実施の形態】
以下、図面を参照して本発明の実施の形態を説明する。理解を容易にするために、以下では具体例として横軸を時系列の実験ケースとし、縦軸を遺伝子の発現量とした発現パターンのグラフを用いて説明する。しかし、本発明の適用は遺伝子の発現パターン、あるいは横軸を時系列の実験ケースとした発現パターンに限られるものではない。
【００２２】
例えば、カエルの発生時に働くある遺伝子に注目しその発現及び発現量を調べる場合、まず、大量のカエルの卵を同時に受精させ、ある一定の時間経過ごとにその遺伝子のコピーの有無と数量を計測する方法がある。本発明においては、このような実験において、一定の時間経過後における卵の状態を時系列的に観察し、それぞれの時間において受精された卵の遺伝子の発現及び発現量を測定することを一つの実験ケース（時系列の実験ケース）と捉える。また、実験ケースは、高熱による刺激の有無や投薬の前後で調べるなど実験環境の違いで比較を行う場合や、細胞分裂の時点や栄養素を与えた時点など細胞活動の時期の違いで比較を行う場合、さらに生物のそれぞれの部位同士で比較を行う場合、異なる生物種の間で比較を行う場合なども一つの実験ケースである。
【００２３】
例えば、時系列の実験ケース以外の例として、横軸を体の部位とするものがある。具体的には横軸に脳、心臓、腎臓などの体の部位を取り、それぞれの部位における遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片、たんぱく質等の発現量を結んで発現パターンを得ることができるが、本発明はこのような発現パターンに対しても適用可能である。あるいは、本発明は、個体の種別による違いを見るために、横軸を、例えばヒト、マウスなどの哺乳類や爬虫類、両生類などの個体種別として得た遺伝子、ＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片、たんぱく質等の発現パターンに対しても適用可能である。更には、横軸を、ある実験環境において個体の種別や体の部位に対して注射、ヒートショック、栄養素の付加などの外部からの影響が起きるだろうと思われる要因を加えたものと、加えないものとした場合、あるいは横軸を生物の体の部位、生物種又は時系列を含めこれらを選択し複合した場合の発現パターン一般に対しても適用できるものである。
【００２４】
酵母やヒトなどの生物において、細胞の分裂、エネルギーの生成などのような細胞内活動が生じるとき、細胞の核内のいくつかの遺伝子はコピーされ、その結果タンパク質が生成

される。つまり、遺伝子が機能し始める。これを遺伝子の発現とよび、遺伝子がコピーされる量を遺伝子の発現量とよぶ。実験ケースごとに遺伝子の発現量を調べると、遺伝子の働きに応じて発現量の増減の変化が現れる。これを本発明では発現パターンという。遺伝子以外にもＤＮＡ、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又はタンパク質などの量も計測することがある。

【００２５】

図１は、本発明の方法に用いるシステムの一例の概略構成図である。このシステムは、一連の細胞のプロセスにおいて遺伝子の発現の度合いを数値化したものを格納している遺伝子発現パターンデータ１４、その発現パターンデータより必要な情報を取り出し解析を行うデータ解析部１６、その結果を視覚化して表示するための処理を行う表示処理部１５、また実際に表示する表示装置１１、本システムへの値の入力や選択の操作を行うためのキーボード１２及びマウス１３等のポインティングデバイス、これらの処理をデータ解析部１６に伝える入力処理部１７を備える。ここで、表示処理部１５、データ解析部１６、入力処理部１７は中央処理装置１０上にある。この中央処理装置１０は、コンピュータとそのプログラムによって具体化される。

【００２６】

図３は、遺伝子発現パターンデータ１４に格納された検索対象遺伝子発現パターンデータの具体的な構造を例示したものである。検索対象となる遺伝子群の発現データを二次元配列Exp[][]で表す。すなわち、遺伝子ID（id）に対応する遺伝子が実験ケース（no）における発現の度合いを数値化したデータをExp[id][no]に格納している。

【００２７】

図４は、ある遺伝子の発現パターンデータと、そのデータに基づいて描かれる発現パターンの対応例を説明する模式図である。実験ケースは時系列になっているとする。例えば、一定時間おきに採取したサンプルに対し、同種のＤＮＡチップを用いハイブリダイゼーション反応を行って得られた発現パターンデータである。

【００２８】

図４（ａ）は、遺伝子１の発現パターンデータ例を表し、図４（ｂ）は遺伝子１の発現パターンをグラフ化して表したものである。このグラフの横軸は時系列の時間（分）であり、縦軸は発現量（単位なし）である。図４（ｂ）の曲線（折れ線グラフ）と、図４（ａ）の配列Exp[][]とは対応している。この曲線を二次元配列Exp[][]で表現する場合、遺伝子１のIDである１が最初の添字に入り、次の添字には実験ケースの順番が入る。そして、この配列がもつ値は発現量となる。図示した例の場合、遺伝子１のｉ分後の発現量のデータはExp[1][i+1]のように表すことができ、例えば２分後における発現量５のデータは実験ケース順では3番目なので、配列Exp[1][3]で表せ、この中身はExp[1][3]=5となる。

【００２９】

図５は、図１に示したシステムにおける表示装置１１の表示画面の一例を示す説明図である。この表示画面は、検索対象遺伝子データが入力され検索対象となる遺伝子の属性を表示する検索対象遺伝子表示ボックス２２、検索対象遺伝子表示ボックス２２に入力された複数の遺伝子の遺伝子発現パターンが重ねて同時に表示される検索対象遺伝子用グラフ表示領域２５、検索の元となる遺伝子データが入力され検索する遺伝子の属性を表示する検索遺伝子表示ボックス２３、検索遺伝子表示ボックス２３に入力された遺伝子の遺伝子発現パターンが表示される検索遺伝子用グラフ表示領域２９を有する。

【００３０】

また、検索遺伝子用グラフ表示領域２９に表示されている遺伝子発現パターンから検索候補となる任意の区間を選択するための選択区間設定枠３０、選択区間設定枠３０で選択された区間の発現パターンに加える変形方法を表示した変形オプションボックス２４、変形オプションボックス２４にて選択された変形方式で変形したパターンを表示する変形オプション表示領域２８が用意されている。この変形オプションを用いることにより、例えばある遺伝子の発現が他の遺伝子の発現を阻害する関係にあるものなどを効率よく発見することができる。選択区間設定枠３０は、例えばマウスポインタで枠をドラッグする等の方

法で、横軸方向の位置、幅を任意に調整することができるようになっている。検索対象遺伝子用グラフ表示領域２５には、例えばマウスポインタでドラッグする等の方法で、ユーザが左右に移動して検索開始点を自由に設定することのできる検索区間設定枠２６が設けられている。

【００３１】

選択区間設定枠３０に表示された部分パターンと類似した発現パターンを有する遺伝子の検索は検索ボタン３２をクリックすることにより開始され、検索結果は類似表３１に類似順に表示される。図示の例では、類似表３１には類似度、遺伝子名、検索区間が表示されるようになっている。類似表３１の上部に置かれた詳細表示ボタン３３をクリックすると、類似表３１に表示された個々の遺伝子の詳しい属性が別ウィンドウに表示されるようになっている。

【００３２】

表示画面上の詳細設定ボタン２１をクリックすると、図６（ａ）に略示するような詳細設定ウインドウ４１が表示される。詳細設定ウインドウ４１には、帯のオプションを設定するための図６（ａ）に示す帯オプション設定タグ４２、帯の使用・不使用を設定するための図６（ｂ）に示す帯使用タグ４３、類似パターン検索に用いる帯の幅を設定するための図６（ｃ）に示す帯の幅設定タグ４４、（非）類似度の算出方法を選択するための図６（ｄ）に示す（非）類似度タグ４５、検索する遺伝子と検索対象となる遺伝子に対し正規化処理の適用・不適用を設定するための図６（ｅ）に示す正規化タグ４６が含まれる。ここで帯とは、後述するように、発現パターンの発現量の値に任意の値を加減してできた、その発現パターンを中心として縦軸方向に幅を持った領域をいう。

【００３３】

ユーザは、図５に示した表示画面において、検索遺伝子用グラフ表示領域２９に表示されている遺伝子発現パターンの中で検索候補となる任意の区間を選択区間設定枠３０によって選択する。この選択区間設定枠３０によって切り取られた曲線（パターン形状）をもとに、検索対象遺伝子用グラフ表示領域２５に表示されている検索対象遺伝子の遺伝子発現パターンの中から類似する曲線を抽出する。すなわち、選択区間設定枠３０によって切り取られたパターン形状に類似する曲線を部分的に含むパターンを検索対象遺伝子用グラフ表示領域２５に表示されている遺伝子発現パターンの中から抽出する。

【００３４】

ユーザは検索区間設定枠２６を左右に動かし、検索対象遺伝子用グラフ表示領域２５における検索開始点を決定する。例えば、検索区間設定枠２６内における遺伝子のカスケードの状況を観測するとき、ユーザが適宜、区間設定枠を移動し、類似している遺伝子を調べればよい。検索開始点は自動的に設定することも可能である。また、複数の検索開始点を自動設定し、検索開始点を順次移動しながら自動検索することも可能である。

【００３５】

図７は、本発明による類似遺伝子発現パターン抽出方法の概略フローを示した図である。まず、図１に示した遺伝子発現パターンデータ１４からデータ解析部１６へデータを読み込む（ステップ１１）。個々の遺伝子発現パターンデータの具体的な形は、先に図３にて説明したようなものである。次に、ユーザは、図１のデータ解析部１６に読み込んだ遺伝子発現パターンデータのうち、検索対象となる遺伝子の発現パターンを取り出す（ステップ１２）。さらに、検索したい遺伝子の発現パターンを取り出す（ステップ１３）。ここでは、後述するように、ユーザ自身が任意に作成したデータを検索する遺伝子発現パターンデータとして入力するなどの方法も可能である。次にユーザは、検索遺伝子発現パターンデータの詳細設定と類似度（非類似度）の距離計算方式を決定する（ステップ１４）。

【００３６】

次に、検索対象遺伝子発現データにおける検索遺伝子の検索開始点を決める（ステップ１５）。ここでは、ユーザが図５の表示画面上で検索区間設定枠２６を左右に動かすことにより検索対象遺伝子発現パターンデータにおける検索開始点の値を決定し、これを変数Ｔｓに保持する。次に、検索対象遺伝子群から任意の検索遺伝子に類似した遺伝子を抽出す

る（ステップ１６）。最後に、抽出された検索対象遺伝子を、類似表３１に類似している順に並べて表示する（ステップ１７）。

【００３７】

図８は、図７のステップ１１における検索対象遺伝子発現パターンデータの登録に関する処理の詳細フローである。

まず、図９に示すように、本システムに蓄積された遺伝子発現パターンデータ１４を別ウィンドウのローカル遺伝子データベース表示ボックス５１に表示する。ローカル遺伝子データベース表示ボックス５１には格納された遺伝子の遺伝子IDと遺伝子名が表示されるので、検索対象とする任意の遺伝子データを例えばポインティングデバイスで検索対象遺伝子表示ボックス５２にドラッグアンドドロップすることにより選択し格納する。検索対象遺伝子表示ボックス５２のデータは、図５に示した検索対象遺伝子表示ボックス２２に入力される。検索対象遺伝子表示ボックス２２に検索対象遺伝子が入力されると、検索対象遺伝子用グラフ表示領域２５に発現度に応じた検索対象遺伝子の遺伝子発現パターンが表示される。同時に、図３にて説明したようにして、id番目の遺伝子のno個目の実験ケースにおける各遺伝子の発現量を二次元配列Exp[id][no]に登録する（ステップ２１）。次に、変数gene_numに登録する遺伝子の総数を保持し、変数case_numに登録する遺伝子の実験ケースの総数を保持する（ステップ２２）。

【００３８】

図１０は、図７のステップ１２に示した検索遺伝子発現パターンデータの入力に関する処理の詳細フローである。

ローカル遺伝子データベースより検索遺伝子を選択するときは、図９のローカル遺伝子データベース表示ボックス５１より検索する任意の遺伝子発現パターンデータをポインティングデバイスで検索遺伝子表示ボックス５３にドラッグアンドドロップすることにより選択する（ステップ３１）。検索遺伝子表示ボックス５３のデータは、図５に図示した検索遺伝子表示ボックス２３に入力される。検索遺伝子表示ボックス２３に検索遺伝子が入力されると、検索遺伝子用グラフ表示領域２９に検索遺伝子の遺伝子発現パターンが表示される。

【００３９】

次に、検索遺伝子発現パターンを配列Target[]にコピーする（ステップ３３）。ローカル遺伝子データベースより検索遺伝子を選択しないとき、ユーザは任意に遺伝子発現パターンを作成し、この作成した遺伝子発現パターンを配列Target[]にコピーする（ステップ３２）。

【００４０】

図１１は、図７のステップ１３に示した検索遺伝子発現パターンデータの設定に関する詳細フローである。

まず、図５の検索遺伝子用グラフ表示領域２９に表示された全区間の検索遺伝子発現パターンの中から、利用者が着目したい区間を選択区間設定枠３０を用いて選択する。選択した区間の開始点の値、終了点の値を保持しておく（ステップ４１）。

【００４１】

図１２は、図５の選択区間設定枠３０の検索範囲に対応する遺伝子発現パターンデータの具体的な構造を示した説明図である。検索する遺伝子の発現パターンデータを一次元配列Target[]で表す。すなわち、遺伝子ID（id）の実験ケースｓ，ｓ+1，…，ｓ+case_num-1におけるデータを検索遺伝子の発現データとするとき、Target[]はTarget[1]=Exp[id][ｓ]，…，Target[case_num]=Exp[id][ｓ+case_num-1]となる。変数case_numはTarget[]に格納する実験ケースの総数である。

【００４２】

次に、検索遺伝子に対して、どのような相互関係を持つ遺伝子を検索対象遺伝子群から抽出したいかにより、選択した区間の検索遺伝子発現パターンを変形する（ステップ４２）。この発現パターンの変形は、変形オプションボックス２４を用いて行われる。変形オプションボックス２４の中にチェックが付されている変形方式が選択された変形方式であり

、選択区間設定枠３０によって囲まれた遺伝子発現パターンの一部は、変形オプション表示領域２８上の変形表示ウインドウに、変形を施さないそのままの形状の曲線とともに、それぞれ選択した変形方式に従って変形されて表示される。変形オプション表示領域２８より任意の変形方法を１つ選び、ポインティングデバイスによりスライドケース２７にドラッグアンドドロップすると、検索対象遺伝子用グラフ表示領域２５に検索区間設定枠２６が設定される。スライドケース２７内で検索区間設定枠２６を左右に動かすことにより検索開始点を自由に設定することができる。

【００４３】

変形オプションとしては、曲線の上下を反転する「抑制型」、発現量の倍率を変え曲線を縦軸方向に拡大又は縮小する、あるいは曲線を横軸方向に拡大あるいは縮小する「倍率変化型」、さらに、ユーザが選択した検索遺伝子発現パターンに対して、変形オプション表示領域上でポインティングデバイス等を用いて任意の変化を加える「ユーザ指定型」などがある。抑制型で変形したパターンを用いて類似検索をすると、検索遺伝子に対する抑制遺伝子を検索することができ、倍率変化型で変形したパターンを用いて類似検索をすると、発現量を増幅する遺伝子などを検索することができる。また、ユーザ指定型の変形パターンを用いると、ユーザの考えを反映したり、実験による誤差を修正して遺伝子検索を行うことができる。変形オプションの種類は、ここに例示したものに限られず、適宜のものを追加することが可能である。

【００４４】

次に、類似度（非類似度）の距離計算方式を決定する（ステップ４３）。類似度あるいは非類似度とは、２つの発現パターンの類似の程度を表す指標である。この指標には、距離のように値の小さい方が類似性が高いことを表す場合と、相関係数のように値の大きい方が類似性が高いことを表す場合とがある。前者の指標を非類似度、後者の指標を類似度という。類似度（非類似度）の距離計算方式は、図６（ｄ）に示した（非）類似度タグ４５によって任意の方法を選択する。図示した例では、類似度としてピアソンの相関係数を選択でき、また、非類似度としてユークリッド平方距離、標準化ユークリッド平方距離、マハラノビスの（汎）距離、ミンコフスキー距離等を選択することができる。ピアソンの相関係数は、図２２（ｆ）に示したような常に一定の倍率を持って発現が見られる遺伝子発現パターンを検索するときに有効であることが知られているので、類似度としてピアソンの相関係数を選択する時は後述する帯を自動的に不使用にするようにしてもよい。

【００４５】

次に、ユーザが図６（ｃ）に示す帯の幅タグ設定４４に所望の値を代入することにより帯の幅を決定する。この値をWとする。（ステップ４４）。帯の幅について、図１３を用いて説明する。帯とは、例えば図１３（ａ）に示すような発現パターンがあったとき、図１３（ｂ）に示すように、任意の値（ここでは１）を発現量の値に加減してできた、その発現パターンを中心とした縦軸方向に幅を持った領域をいう。帯の幅を小さくすれば、選択区間設定枠３０によって囲まれた検索遺伝子発現パターンの一部と類似度の高い発現パターンを有する遺伝子のみが検索され、逆に帯の幅を大きくすれば、類似度の比較的低い発現パターンを有する遺伝子も検索されることになる。

【００４６】

また、ある遺伝子に着目し、その遺伝子発現パターンを異なる体の部位の間で比べる時や、投薬前、投薬後で比べる時などでは、個々の実験で温度やチップ上のスポッティングの量などの実験環境の違いから、測定された遺伝子発現データが均一でないときがある。すなわち実験によって、遺伝子全体でみた時、発現量の平均値や分散値に違いがあることがある。この時、正規化処理を行うとこれらの違いを補正することができる。本システムでは正規化処理を適用するか否かを選択できるようにしている（ステップ４５）。正規化処理の使用に関しては、図６（ｅ）に示した正規化タグ４６の選択により決定する。正規化を選択した場合、検索遺伝子、検索対象遺伝子ともに発現パターンデータに対し正規化処理を行い、その値をもとに図５に示した表示画面の検索対象遺伝子用グラフ表示領域２５及び検索遺伝子用グラフ表示領域２９に表示する。

【００４７】
図１４は、図７に示したステップ１６の類似パターン抽出処理の詳細フローである。まず、帯を用いて発現パターンが全区間内において帯の幅の中に収まっている類似パターンを検索する。このとき、横軸上の開始点をＴｓとする（ステップ５１）。検索の途中で帯内に入り類似パターンと認められた曲線はその時点でメモリに保存してもよいし、検索がすべて終了した後でメモリに保存してもかまわない。
【００４８】
図１５は、帯を用いたことによって類似パターンとして認められるパターンと、類似パターンとは認められないパターンの例を示す説明図である。類似パターンと認められるのは、発現パターンが選択された全区間内において帯の幅の中に収まっているパターンであり、これには図１５（ａ）に示すように遺伝子発現パターンが帯の変化とほぼ同じである場合、図１５（ｂ）に示すように遺伝子発現パターンは帯の中央を通らないが全ての値は帯と重なっている場合、図１５（ｃ）に示すように遺伝子発現パターンは帯の中央を通らないが全ての値は帯と重なっている場合などがある。
【００４９】
これらの場合は具体的には、遺伝子IDがｉの遺伝子に対する発現データExp[i][Ts], Exp[i][Ts+1], …, Exp[i][Ts+case_num-1]と検索する遺伝子の発現パターンデータTarget[1], Target[2], …, Target[case_num]との間に、次の〔数１〕の関係が全て成り立つ場合であり、〔数１〕の関係を満たす遺伝子ｉは検索遺伝子と類似の発現パターンを有するとみなす。
【００５０】
【数１】
Target[1]-W≦Exp[i][Ts]≦Target[1]+W,
Target[2]-W≦Exp[i][Ts+1]≦Target[2]+W,
……,
Target[case_num]-W≦Exp[i][Ts+case_num-1]≦Target[case_num]+W
【００５１】
一方、類似パターンと認められないのは、図１５（ｄ）に示すように遺伝子発現パターンのうち初期値から中ほどの値にかけて帯と重なるが最後ははみ出てしまう場合、図１５（ｅ）に示すように初期の値が帯からはみ出てしまう場合、図１５（ｆ）に示すように初期の値と最後の値は帯と重なるが途中ではみ出てしまう場合などである。
【００５２】
本発明では、図１６に略示するように、横軸区間は固定したまま、（ａ）～（ｄ）のように帯を縦軸に関して上下にスライドさせながら、帯の範囲に含まれる遺伝子発現パターンを抽出する。ここで、検索遺伝子の発現パターンと類似パターンと認められるのは、図１５（ａ）のように遺伝子発現パターンが帯の変化とほぼ同じ場合、図１５（ｂ），（ｃ）のように遺伝子発現パターンは帯の中央を通らないが全ての値は帯と重なっている場合である。
【００５３】
しかし、上記の方法によって発現パターンが類似しているとは認められない遺伝子の中にも相互の発現が関連性を有するものが含まれていることがある。そこで、検索区間の前後の値が帯からはみ出ていても、ある一定の割合で中央の値が重なっていれば類似パターンと認めるオプションを設ける。このオプションの設定は図６（ａ）に示した帯オプション設定タグ４２で行う。帯オプションを例えば８０％に指定するということは、選択区間設定枠３０の横軸方向の幅をその中央値を中心として８０％の範囲に縮小したのと同じ効果を有する。したがって、図示の例の場合、図１７に示すように、帯オプション無しの検索では類似パターンとは認められなかった図１５（ｄ）や図１５（ｅ）のようなパターンも類似パターンとして認められることになる。ただし、図１５（ｆ）のように中央の値が帯からはみ出ている場合は、帯オプションを設定したとしても類似パターンとは認められない。図１６（ｄ）に示すように、帯が区間設定枠の上部に達した場合、帯による検索は終

了する。

【００５４】

図１４に戻って、次にメモリに保存した帯内に入る曲線に対し（非）類似度を算出する。すなわち、図１１のステップ４３で設定した距離計算法に基づいて、検索遺伝子の発現パターンTarget[1], Target[2], …, Target[case_num]と検索対象遺伝子の発現パターンExp[i][Ts], Exp[i][Ts+1], …, Exp[i][Ts+case_num-1]との間の（非）類似度を求める（ステップ５２）（ｉは、帯に入る遺伝子発現パターンの遺伝子ID）。次に、類似性の高い順（（非）類似度の高い順）に並べ替え（ステップ５３）、表示画面の類似表３１に表示する。

【００５５】

図１８は、詳細表示ボタン３３をクリックしたとき表示される類似度の降順による遺伝子データの表示例である。「類似度」の欄６２には、（非）類似度タグ４５で選択した類似度あるいは非類似度の評価方法によって評価された類似度あるいは非類似度の数値が表示される。「検索区間」の欄６３には、検索区間における検索対象データ６０と検索データ６１の曲線データが表示される。また、「遺伝子に関する詳細表示」の欄６４には、遺伝子が発現している生物体の部位、遺伝子の塩基配列等、その遺伝子に関連する情報が表示される。

ユーザーは図５に示した表示画面上で検索区間設定枠２６を横軸方向（左右）に動かすことにより検索開始点の設定を行い検索を行うが、ユーザの検索したい区間において、検索開始点を自動的に移動しながら検索を行うようにすることも可能である。

【００５６】

図１９は、検索開始点を自動的に移動しながら検索を行う場合の説明図である。この場合には、まず図１９（ａ）に示すように、検索対象遺伝子用グラフ表示領域２５上の検索区間設定枠７１をユーザの検索したい区間をカバーするように左右方向に広げる。次に、検索開始点のステップ移動の距離を決める。これらの設定をしておくことにより、図１９（ｂ）に略示するように、検索区間設定枠７１内で検索開始点が横軸方向に▲１▼, ▲２▼, …のように一定の間隔で自動的に設定される。そして、各検索開始点において上方向に移動する帯が検索区間設定枠７１の上部に達すると、設定された移動距離だけ帯を横軸方向にずらして再び帯を上方向に移動しながら検索を反復することで、検索区間設定枠７１内の全域に亘って自動的に検索を行なうことができる。

【００５７】

更に、帯または曲線を横にスライドしながら検索することも可能である。図２０は、帯又は曲線のスライド方向を説明する図である。図２０（ａ）は検索区間設定枠７３を縦軸方向に長く設定した場合の説明図であり、前述のように、検索遺伝子の発現パターンの一部あるいはそれを元に形成された帯を検索区間設定枠７３内を上下に移動して類似パターンの検索を行う。このとき、図中に白抜き矢印で示すように、検索区間設定枠７３は横軸方向に自由に移動することが可能である。

【００５８】

これに対して図２０（ｂ）は、検索区間設定枠７４を横軸方向に長く設定した場合の説明図であり、検索遺伝子の発現パターンの一部あるいはそれを元に形成された帯を横軸方向にスライドさせながら検索を行う。図２１は発現パターンの一部あるいはそれを元に作成された帯を横軸方向にスライドさせながら検索を行う様子を模式的に示したものであり、図２１（ａ）～（ｄ）に示すように、発現量区間を固定したまま、帯を横軸に関して左から右へスライドさせ、帯の範囲に含まれる遺伝子曲線を抽出する。この方法は遺伝子のカスケードの検出に有効である。この場合にも、図中に白抜き矢印で示すように、検索区間設定枠７４を縦軸方向に自由に移動することが可能である。

【００５９】

以上、代表的な例を用いて本発明を具体的に説明したが、遺伝子発現データベースはローカルなものだけではなく、ネットワーク上のものも利用できる。また、　一つの検索遺伝子について、複数の変形オプションを同時に選択し、これらを用い並行処理により同時に

検索することも可能である。更に、上記説明では検索する遺伝子は一度に１つしか選択しなかったが、複数選択して並行処理により各遺伝子の発現パターンにそれぞれ類似する曲線を同時に検索することも可能である。

【００６０】

【発明の効果】

以上説明したように、本発明によれば、遺伝子等の大量の未知の生体高分子群から既知の生体高分子の発現情報をもとに関連ある生体高分子を抽出し、その生体高分子の生物学的機能を効率よく推測することができる。

【図面の簡単な説明】

【図１】本発明の方法に用いるシステムの一例の概略構成図。

【図２】本発明による類似パターン検索の例を示す図。

【図３】検索対象遺伝子発現パターンデータの例を示す図。

【図４】グラフ上の遺伝子発現パターンデータと二次元配列における遺伝子発現パターンデータの対応例を示す図。

【図５】表示装置の表示画面の一例を示す説明図。

【図６】詳細設定ウインドウの説明図。

【図７】本発明による類似遺伝子発現パターン抽出処理の概要を示すフローチャート。

【図８】検索対象遺伝子発現パターンデータの登録処理に関するフローチャート。

【図９】遺伝子データから検索対象遺伝子と検索遺伝子を選択する例を示す説明図。

【図１０】検索遺伝子発現パターンデータの入力処理に関するフローチャート。

【図１１】検索遺伝子発現パターンの詳細設定及び距離計算法に関するフローチャート。

【図１２】検索遺伝子発現パターンデータの例を示す図。

【図１３】帯の幅に関する説明図。

【図１４】類似パターン抽出処理の詳細設定に関するフローチャート。

【図１５】類似パターン抽出処理において、類似パターンと認められる場合と認められない場合の例を示す図。

【図１６】区間設定枠内の帯の縦軸方向のスライドを説明する図。

【図１７】帯オプション設定例を示す図。

【図１８】類似性の高い順による遺伝子データの表示例を示す図。

【図１９】検索区間を指定し自動的に検索する例を示す図。

【図２０】検索区間設定の説明図。

【図２１】区間設定枠内の帯の横軸方向のスライドに関する例を示す図。

【図２２】従来の方法で検索できなかった遺伝子発現パターンの例を示す図。

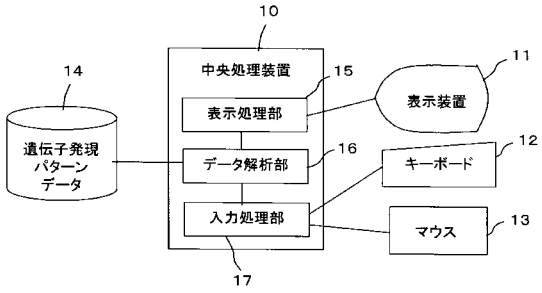
【図２３】遺伝子のカスケードの例を示す図。

【図２４】従来の類似パターン抽出処理法の説明図。

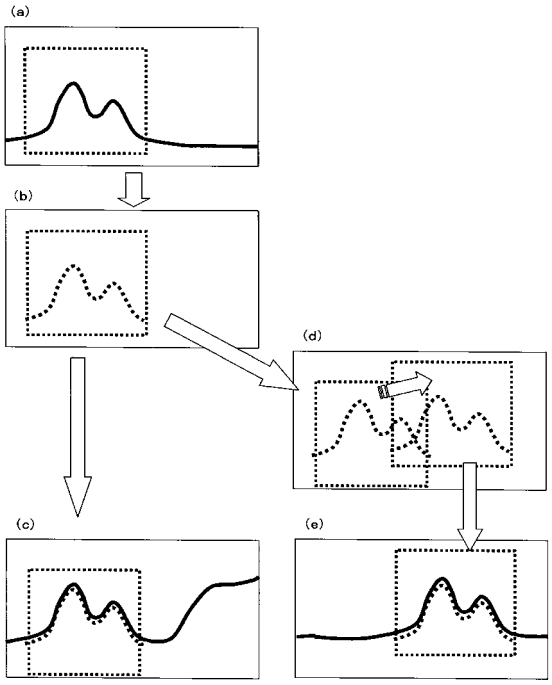
【符号の説明】

１０…中央処理装置、１１…表示装置、１２…キーボード、１３…マウス、１４…遺伝子発現パターンデータ、１５…表示処理部、１６…データ解析部、１７…入力処理部、２１…詳細設定ボタン、２２…検索対象遺伝子表示ボックス、２３…検索遺伝子表示ボックス、２４…変形オプションボックス、２５…検索対象遺伝子用グラフ表示領域、２６…検索区間設定枠、２７…スライドケース、２８…変形オプション表示領域、２９…検索遺伝子用グラフ表示領域、３０…選択区間設定枠、３１…類似表、３２…検索ボタン、３３…詳細表示ボタン、４１…詳細設定ウインドウ、４２…帯オプション設定タグ、４３…帯使用タグ、４４…帯の幅設定タグ、４５…（非）類似度タグ、４６…正規化タグ、５１…ローカル遺伝子データベース表示ボックス、５２…検索対象遺伝子表示ボックス、５３…検索遺伝子表示ボックス、６０…検索対象データ、６１…検索データ、７１…検索区間設定枠、７３…検索区間設定枠、７４…検索区間設定枠

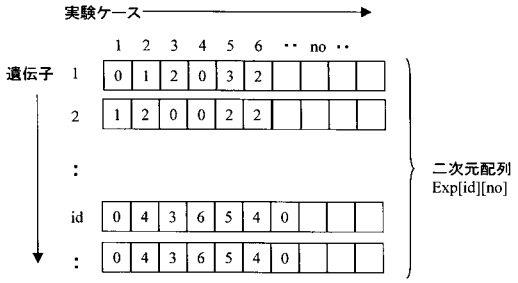
【図１】



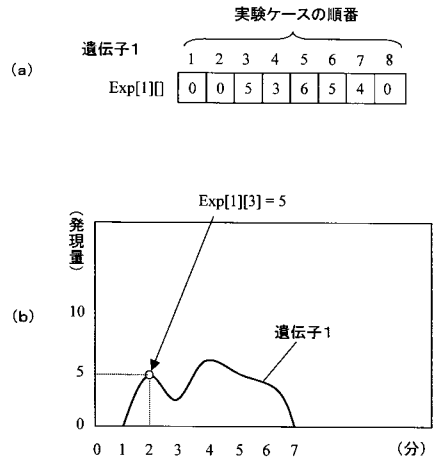
【図２】



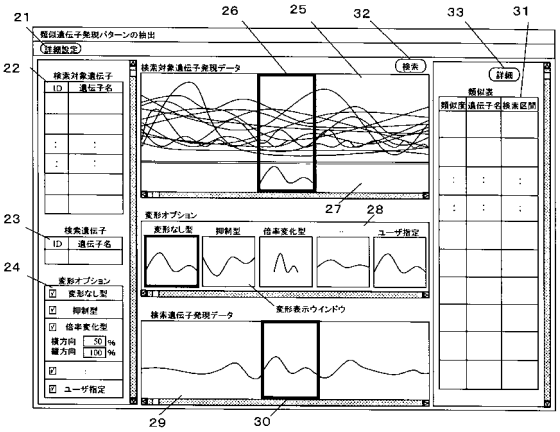
【図３】



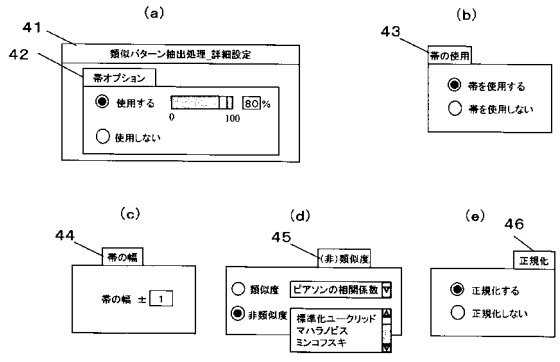
【図４】



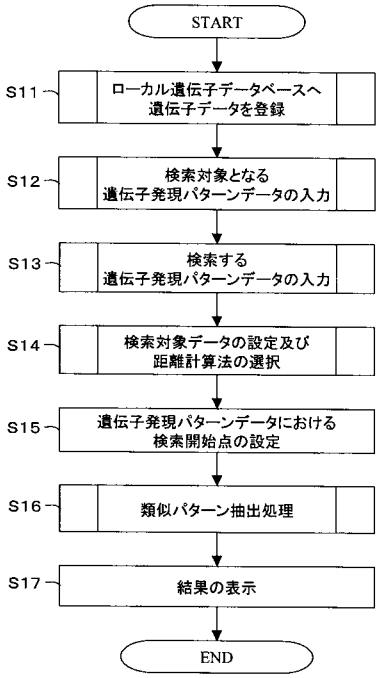
【図５】



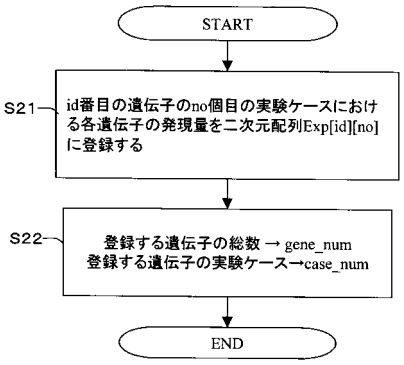
【図６】



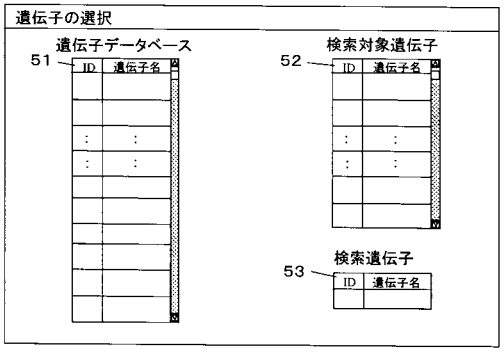
【図７】


START
S11
ローカル遺伝子データベースへ
遺伝子データを登録
S12
検索対象となる
遺伝子発現パターンデータの入力
S13
検索する
遺伝子発現パターンデータの入力
S14
検索対象データの設定及び
距離計算法の選択
S15
遺伝子発現パターンデータにおける
検索開始点の設定
S16
類似パターン抽出処理
S17
結果の表示
END

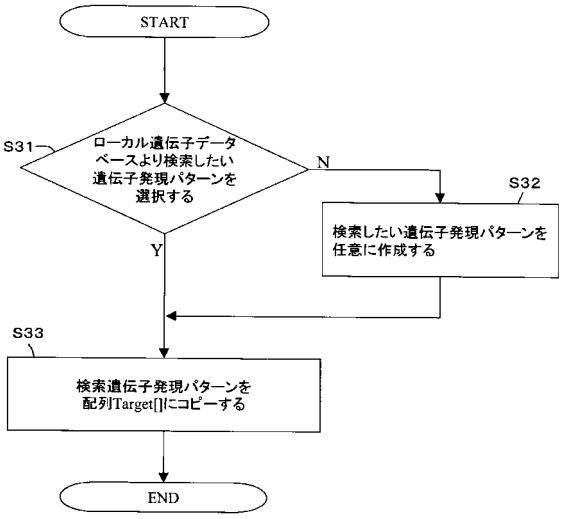
【図８】


START
S21
id番目の遺伝子のno個目の実験ケースにおける各遺伝子の発現量を二次元配列Exp[id][no]に登録する
S22
登録する遺伝子の総数 → gene_num
登録する遺伝子の実験ケース→case_num
END

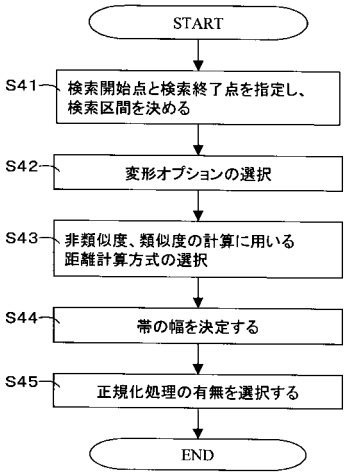
【図９】


遺伝子の選択
遺伝子データベース
51
ID
遺伝子名
検索対象遺伝子
52
ID
遺伝子名
検索遺伝子
53
ID
遺伝子名

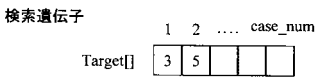
【図１０】


START
S31
ローカル遺伝子データベースより検索したい遺伝子発現パターンを選択する
N
S32
検索したい遺伝子発現パターンを任意に作成する
Y
S33
検索遺伝子発現パターンを配列Target[]にコピーする
END

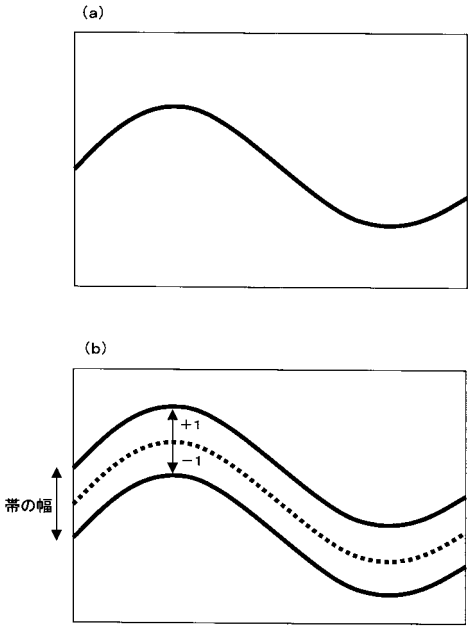
【図１１】



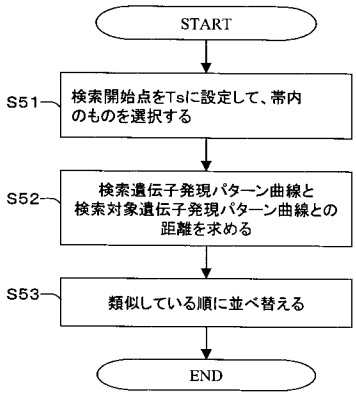
【図１２】



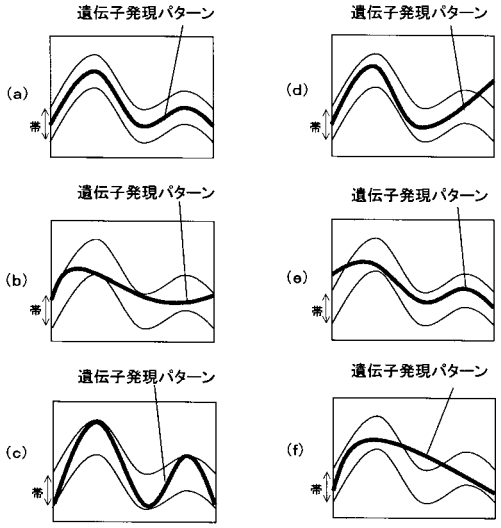
【図１３】



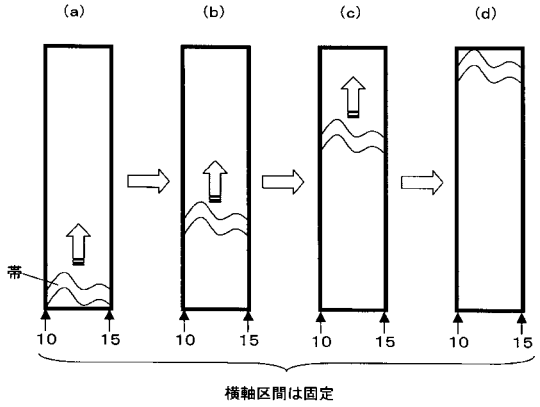
【図１４】



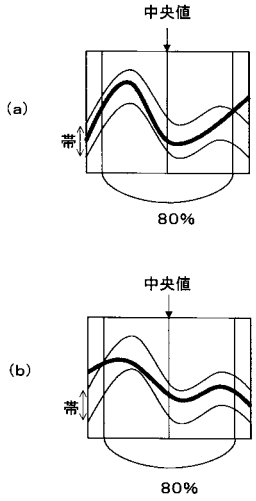
【図１５】



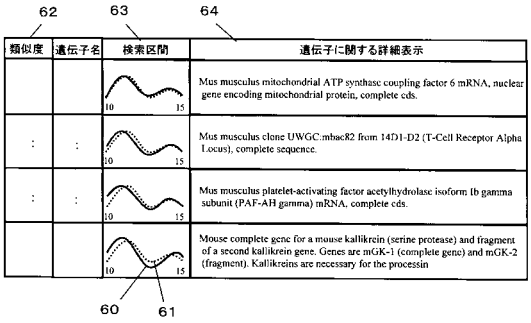
【図１６】



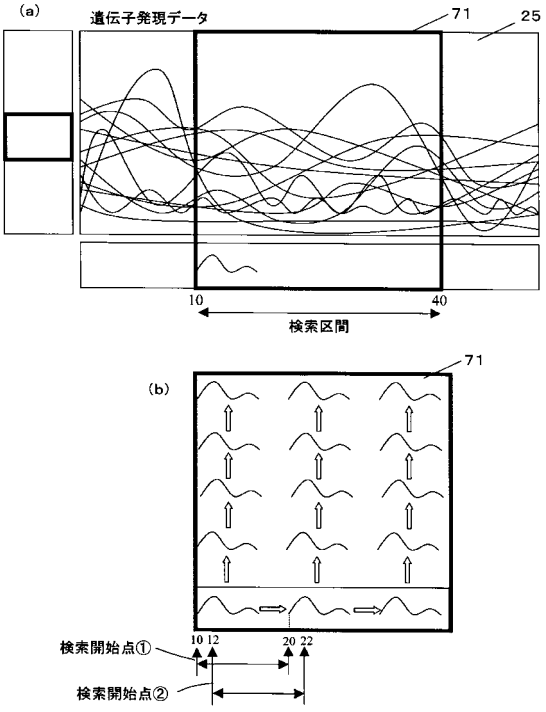
【図１７】



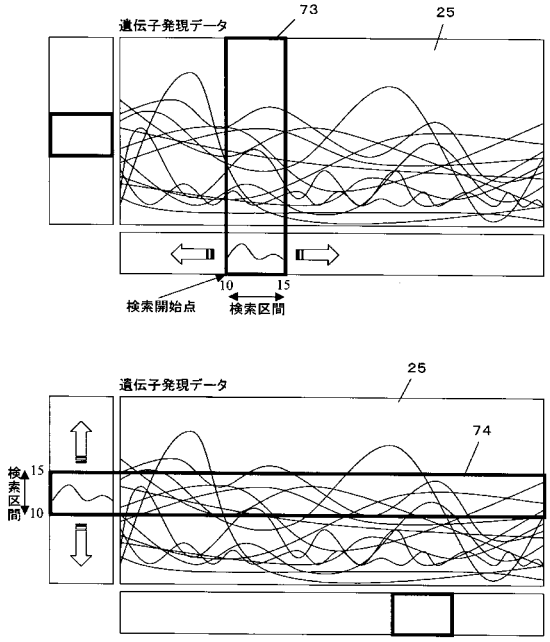
【図１８】



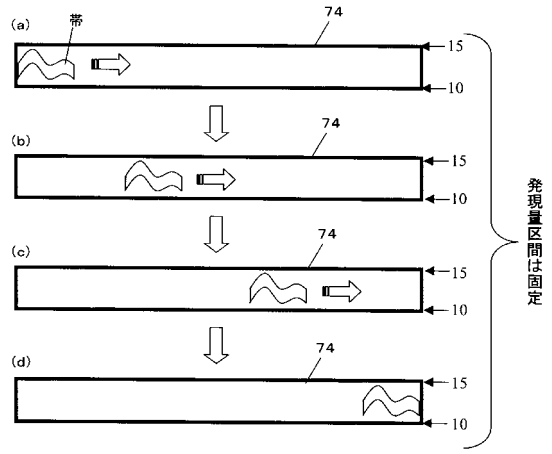
【図１９】



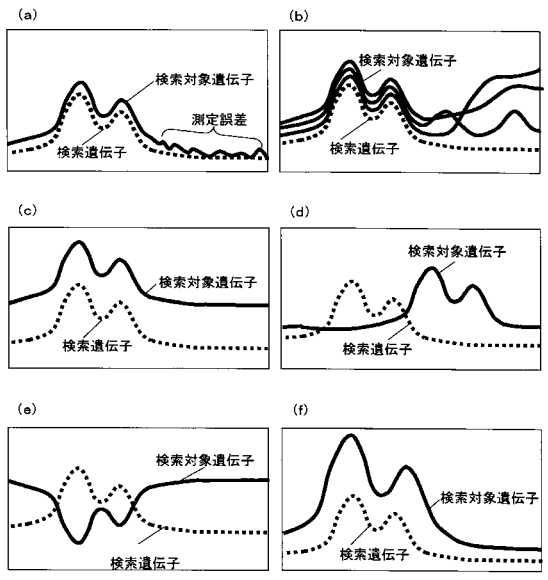
【図２０】



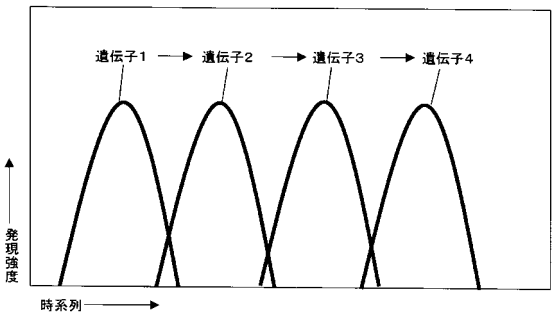
【図２１】



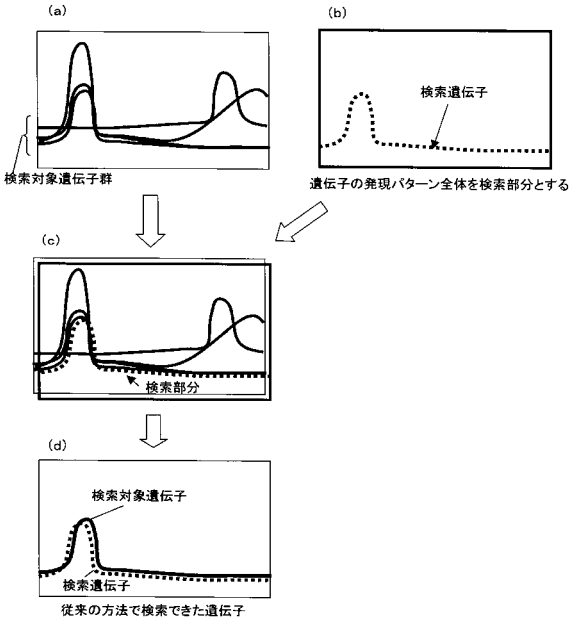
【図２２】



【図２３】



【図２４】


(a)
(b)
検索遺伝子
検索対象遺伝子群
遺伝子の発現パターン全体を検索部分とする
(c)
検索部分
(d)
検索対象遺伝子
検索遺伝子
従来の方法で検索できた遺伝子

フロントページの続き

(72)発明者　野崎　康行
神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内
(72)発明者　中重　亮
神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内
(72)発明者　田村　卓郎
神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

合議体
審判長　赤穂　隆雄
審判官　松田　直也
審判官　山本　穂積

(56)参考文献　国際公開第９９／５８７２０（WO，Ａ１）
特開平７－９８３３３（ＪＰ，Ａ）
特開平８－１２９０３８（ＪＰ，Ａ）
特開平１０－１６０４１９（ＪＰ，Ａ）
特開平９－６８９９４（ＪＰ，Ａ）
江口至洋，“マイクロアレイが知りたい！　第5回　ＤＮＡマイクロアレイ・インフォマティクスの現況と課題”，実験医学，株式会社羊土社，１９９９年9月1日，第１７巻，第１３号，ｐ．１６７０～１６７３

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F19/00
G06Q10/00-50/00