

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5344670号
(P5344670)

(45) 発行日 平成25年11月20日(2013.11.20)

(24) 登録日 平成25年8月23日(2013.8.23)

(51) Int.Cl.
G06F 19/20 (2011.01)

F I
G06F 19/20

請求項の数 15 (全 25 頁)

(21) 出願番号	特願2008-32466 (P2008-32466)	(73) 特許権者	301032942
(22) 出願日	平成20年2月13日 (2008.2.13)		独立行政法人放射線医学総合研究所
(65) 公開番号	特開2009-193273 (P2009-193273A)		千葉県千葉市稲毛区穴川四丁目9番1号
(43) 公開日	平成21年8月27日 (2009.8.27)	(74) 代理人	100064414
審査請求日	平成22年10月22日 (2010.10.22)		弁理士 磯野 道造
		(74) 代理人	100111545
			弁理士 多田 悦夫
		(72) 発明者	笠間 康次
			千葉県千葉市稲毛区穴川四丁目9番1号
			独立行政法人放射線医学総合研究所内
		(72) 発明者	安倍 真澄
			千葉県千葉市稲毛区穴川四丁目9番1号
			独立行政法人放射線医学総合研究所内
		審査官	宮久保 博幸

最終頁に続く

(54) 【発明の名称】 遺伝子発現解析方法、遺伝子発現解析装置、および遺伝子発現解析プログラム

(57) 【特許請求の範囲】

【請求項1】

発現している複数の遺伝子転写産物に由来する複数のピークを一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析方法であって、

複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られるDNAフラグメントの塩基数相当値の位置と、その位置における前記遺伝子転写産物の転写量相当の検出量と、に基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ前記遺伝子転写産物の転写産物種情報として、前記第一の波形中の所定のピークと、そのピークが由来する転写産物種と、を同定して記憶した参照プロファイルデータを予め取得しておく参照プロファイルデータ取得工程と、

10

複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られる測定対象物となるDNAフラグメントの塩基数相当値の位置と、その位置における前記測定対象物の転写量相当の検出量と、に基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータを作成する測定プロファイルデータ作成工程と、

前記第一の波形および前記第二の波形のうち少なくとも一方の一部または全部の領域を補正しピークの位置を調整する補正処理を行い、前記第二の波形中の着目する領域を含む複数のピークと、前記第一の波形における複数のピークと、を対応付けることで、前記着目する領域を含む複数のピークと、当該複数のピークに相当する第一の波形中の複数のピークと、を対応付けるピーク対応付け工程と、

対応付けされた前記測定プロファイルデータのピークの由来する遺伝子転写産物の情報

20

を前記転写産物種情報から読み取り、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解析する遺伝子発現解析工程と

、
を含み、

前記ピーク対応付け工程は、

前記参照プロファイルデータ取得工程で取得された前記参照プロファイルデータのピークの位置と、前記測定プロファイルデータ作成工程で作成された前記測定プロファイルデータのピークの位置と、が一致する場合は、一致する前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付け、

前記参照プロファイルデータ取得工程で取得された前記参照プロファイルデータのピークの位置と、前記測定プロファイルデータ作成工程で作成された前記測定プロファイルデータのピークの位置と、が一部または全部ずれている場合は、これらのうちの少なくとも一方の波形について、これらの波形の類似度が最も高くなるように一部または全部の領域を補正しピークの位置を調整する補正処理を行った上で、前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付ける

10

ことを特徴とする遺伝子発現解析方法。

【請求項 2】

前記参照プロファイルデータ取得工程は、

前記参照プロファイルデータを、

既知のプロファイルデータを保存しているデータベースから取得するか、

前記転写産物種情報から人工的に作成して取得するか、

既知のプロファイルデータ若しくは前記測定プロファイルデータに1つ以上のピークを追加或いは削除することによって取得するか、

前記参照プロファイルデータを複数用いて合成することによって取得するか、または、

前記測定プロファイルデータを複数用いて合成することによって取得する

ことを特徴とする請求項 1 に記載の遺伝子発現解析方法。

20

【請求項 3】

前記ピーク対応付け工程における補正処理が、ガウス関数に基づく関数近似によって行われることを特徴とする請求項 1 または請求項 2 に記載の遺伝子発現解析方法。

【請求項 4】

前記ピーク対応付け工程における補正処理が、

前記参照プロファイルデータのピークの位置を基準として前記測定プロファイルデータのピークの位置を移動させるか、

前記測定プロファイルデータのピークの位置を基準として前記参照プロファイルデータのピークの位置を移動させるか、または、

前記参照プロファイルデータのピークの位置と前記測定プロファイルデータのピークの位置の双方を移動させて、

前記測定プロファイルデータのピークと前記参照プロファイルデータのピークを対応付ける

ことを特徴とする請求項 1 から請求項 3 のいずれか 1 項に記載の遺伝子発現解析方法。

30

40

【請求項 5】

前記遺伝子発現解析工程は、

前記参照プロファイルデータのピークと前記測定プロファイルデータのピークの対応付けができたピークと、対応付けができなかったピークと、が区別できるように表示するとともに、前記参照プロファイルデータのピークに、当該ピークの由来となる遺伝子に関する遺伝子情報が付加されている場合は、当該遺伝子情報を引用することにより前記測定プロファイルデータにおいて対応付けされたピークの遺伝子を特定し、遺伝子の発現状態を解析する

ことを特徴とする請求項 1 から請求項 4 のいずれか 1 項に記載の遺伝子発現解析方法。

【請求項 6】

50

前記遺伝子発現解析工程には、

前記ピーク対応付け工程で対応付けができなかったピークについて、当該ピークに関する関連情報を付加する工程が含まれている

ことを特徴とする請求項 1 から請求項 5 のいずれか 1 項に記載の遺伝子発現解析方法。

【請求項 7】

前記関連情報が、前記波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報のうち少なくとも 1 つを含んでいる

ことを特徴とする請求項 6 に記載の遺伝子発現解析方法。

【請求項 8】

発現している複数の遺伝子転写産物に由来する複数のピークを一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析装置であって、

複数の遺伝子転写産物を逆転写した cDNA の一部を増幅して得られる DNA フラグメントの塩基数相当値の位置と、その位置における前記遺伝子転写産物の転写量相当の検出量と、に基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ前記遺伝子転写産物の転写産物種情報として、前記第一の波形中の所定のピークと、そのピークが由来する転写産物種と、を同定して記憶した参照プロファイルデータを予め取得しておく参照プロファイルデータ取得手段と、

複数の遺伝子転写産物を逆転写した cDNA の一部を増幅して得られる測定対象物となる DNA フラグメントの塩基数相当値の位置と、その位置における前記測定対象物の転写量相当の検出量と、に基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータを作成する測定プロファイルデータ作成手段と、

前記第一の波形および前記第二の波形のうち少なくとも一方の一部または全部の領域を補正しピーク的位置を調整する補正処理を行い、前記第二の波形中の着目する領域を含む複数のピークと、前記第一の波形における複数のピークと、を対応付けることで、前記着目する領域を含む複数のピークと、当該複数のピークに相当する第一の波形中の複数のピークと、を対応付けるピーク対応付け手段と、

対応付けされた前記測定プロファイルデータのピークの由来する遺伝子転写産物の情報を前記転写産物種情報から読み取り、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解析する遺伝子発現解析手段と

を有し、

前記ピーク対応付け手段は、

前記参照プロファイルデータ取得手段で取得された前記参照プロファイルデータのピーク的位置と、前記測定プロファイルデータ作成手段で作成された前記測定プロファイルデータのピーク的位置と、が一致する場合は、一致する前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付け、

前記参照プロファイルデータ取得手段で取得された前記参照プロファイルデータのピーク的位置と、前記測定プロファイルデータ作成手段で作成された前記測定プロファイルデータのピーク的位置と、が一部または全部ずれている場合は、これらのうちの少なくとも一方の波形について、これらの波形の類似度が最も高くなるように一部または全部の領域を補正しピーク的位置を調整する補正処理を行った上で、前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付ける

ことを特徴とする遺伝子発現解析装置。

【請求項 9】

前記参照プロファイルデータ取得手段は、

前記参照プロファイルデータを、

既知のプロファイルデータを保存しているデータベースから取得するか、

前記転写産物種情報から人工的に作成して取得するか、

既知のプロファイルデータ若しくは前記測定プロファイルデータに 1 つ以上のピークを

10

20

30

40

50

追加或いは削除することによって取得するか、

前記参照プロファイルデータを複数用いて合成することによって取得するか、または、前記測定プロファイルデータを複数用いて合成することによって取得することを特徴とする請求項8に記載の遺伝子発現解析装置。

【請求項10】

前記ピーク対応付け手段における補正処理が、ガウス関数に基づく関数近似によって行われることを特徴とする請求項8または請求項9に記載の遺伝子発現解析装置。

【請求項11】

前記ピーク対応付け手段における補正処理が、前記参照プロファイルデータのピークの位置を基準として前記測定プロファイルデータのピークの位置を移動させるか、前記測定プロファイルデータのピークの位置を基準として前記参照プロファイルデータのピークの位置を移動させるか、または、前記参照プロファイルデータのピークの位置と前記測定プロファイルデータのピークの位置の双方を移動させて、前記測定プロファイルデータのピークと前記参照プロファイルデータのピークを対応付ける

ことを特徴とする請求項8から請求項10のいずれか1項に記載の遺伝子発現解析装置。

【請求項12】

前記遺伝子発現解析手段は、前記参照プロファイルデータのピークと前記測定プロファイルデータのピークの対応付けができたピークと、対応付けができなかったピークと、が区別できるように表示するとともに、前記参照プロファイルデータのピークに、当該ピークの由来となる遺伝子に関する遺伝子情報が付加されている場合は、当該遺伝子情報を引用することにより前記測定プロファイルデータにおいて対応付けされたピークの遺伝子を特定し、遺伝子の発現状態を解析する

ことを特徴とする請求項8から請求項11のいずれか1項に記載の遺伝子発現解析装置。

【請求項13】

前記遺伝子発現解析手段には、前記ピーク対応付け手段で対応付けができなかったピークについて、当該ピークに関する関連情報を付加する手段が含まれている

ことを特徴とする請求項8から請求項12のいずれか1項に記載の遺伝子発現解析装置。

【請求項14】

前記関連情報が、前記波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報のうち少なくとも1つを含んでいる

ことを特徴とする請求項13に記載の遺伝子発現解析装置。

【請求項15】

請求項1から請求項7に記載の遺伝子発現解析方法をコンピュータに実行させる

ことを特徴とする遺伝子発現解析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、発現している複数の遺伝子転写産物に由来する複数のピークを有する一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析方法、遺伝子発現解析装置、および遺伝子発現解析プログラムに関する。

【背景技術】

10

20

30

40

50

【 0 0 0 2 】

ゲノム科学は「ヒトの遺伝子情報（ヒトゲノム）の解読完了」という象徴的な事件の後、遺伝子発現に関する制御メカニズムの解明、遺伝子産物の機能解明という、いわゆるポストシーケンス研究へと移行しつつある。ポストシーケンス研究を進めると、さまざまな生命現象を解明することができるので学術的に意義が高いだけでなく、医薬品の開発にも多大な貢献をもたらす、高度なオーダーメイド治療等を実現できる可能性がある。そのため、ポストシーケンス研究の進展に対する期待度も非常に大きい。

【 0 0 0 3 】

このようなポストシーケンス研究の一例として、遺伝子の発現状態を解析する遺伝子発現解析が挙げられる。同じ遺伝子でも、生体の状態等によって時々刻々とその発現量に変化しており、オルタナティブスプライシング等によって一つの遺伝子から異なるタンパク質に対応した複数の mRNA が生成されている。また、高等動物や植物では、タンパク質に翻訳されないノンコーディング RNA やマイクロ RNA も生成されており、これらが遺伝子発現制御を行っていることや、様々な生物種で保存されていることなども明らかになりつつある。

10

これらの RNA の中でもタンパク質の発現に直接的に関係のある mRNA の発現量を解析して遺伝子発現解析を行うことは、前記したように、より高度なオーダーメイド治療等を実現するためにも特に重要である。

【 0 0 0 4 】

遺伝子発現解析を行うための手法としては、遺伝子の発現状態を網羅的に解析することができるディファレンシャルディスプレイ法や S A G E (Serial Analysis of Gene Expression) 法、DNA マイクロアレイ法、DNA チップ法などが広く用いられている。

20

【 0 0 0 5 】

また、近年、遺伝子発現解析を行うための手法として、網羅的かつより高精度な遺伝子発現解析を可能にした高カバー率遺伝子発現解析法 (High Coverage Expression Profiling 法 (以下、「HiCEP 法」と称する。)) が開発され、注目を浴びている (例えば、特許文献 1 参照)。

HiCEP 法は、mRNA を逆転写して得られた cDNA を 2 種類の制限酵素により切断し、これに特殊な塩基配列を有するアダプターと称する 20 塩基ほどの DNA フラグメントをライゲーションさせ、さらに前記した特殊な塩基配列を有するアダプターと相補的な塩基配列を有する、蛍光標識された選択的 PCR 用プライマーを用いて選択的 PCR を行い、キャピラリー電気泳動によって種々の長さに応じて分離し、これを解析するキャピラリー DNA シーケンサーを利用して、複数のピークを有する一つの波形として表された測定プロファイルデータを得るものである。このようにすると、前記した複数のピークは、一つの波形に、約 200 程度のピークを有することになる。これは、同じサンプルであれば条件の相違により強度が変わっても原則として同じ泳動位置 (プロファイル上のピークの位置) に、同じ遺伝子の転写産物に由来するピークが検出される。

30

【 0 0 0 6 】

したがって、HiCEP 法は、塩基配列が決定されていない未知遺伝子に対してもその発現状態を解析することができるという利点を有しており、発現している遺伝子の全転写産物 (全 mRNA) に対して解析される転写産物 (mRNA) の割合をカバー率と定義するならば、前述した従来法のカバー率が 10 ~ 30 % であるのに対し、HiCEP 法は 70 ~ 80 % のカバー率を達成している。さらに、± 約 20 % の微小な変動量を確実に捉えることが可能である。このように、HiCEP 法は、従来の DNA マイクロアレイ法等では実現し得なかった高精度・高感度を達成している。

40

【 0 0 0 7 】

【特許文献 1】国際公開第 02 / 048352 号パンフレット

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 8 】

50

しかしながら、ディファレンシャルディスプレイ法やSAGE (Serial Analysis of Gene Expression) 法、DNAマイクロアレイ法、DNAチップ法などは、塩基配列が予め分かっている遺伝子にしか対応できないこと、感度が低い(例えば、検出のために必要なmRNAの変動量は2~3倍必要といわれている。)こと、発現量が大きく変動するものでないとその結果の再現性が十分とはいえないことなどの問題がある。

【0009】

また、特許文献1に記載のHiCEP法は、従来法では成し得なかった高カバレッジ、高精度、高感度で遺伝子の転写産物の発現状態を解析することができるが、電気泳動を使用して分離するため、cDNAフラグメントのサイズが同じ場合や、電気泳動時のcDNAフラグメントの立体的分子構造が原因で、本来の塩基数とは異なる位置にピークがずれて出現してしまうことがある。また、キャピラリーDNAシーケンサーに用いられる電気泳動用のポリマーの劣化や、異なる製造ロットのポリマーの使用、或いは電気泳動時の電圧の変化や温度の変化によって得られる測定プロファイルデータのピークの位置がずれてしまうことがある。また、異なるモデルのシーケンサーを用いる場合は、ポリマーが異なるなど測定条件の差異により、数bpピークの位置が異なって測定されることがあり、異なる研究者のデータを比較する場合、注意が必要である。

10

【0010】

そのため、作成された測定プロファイルデータとリファレンスとなる参照プロファイルデータとを対比させたり、別の施設や別の日に同じ条件で測定されて作成された測定プロファイルデータ同士を対比させたりした場合にピークの位置が一致せず、対比が困難となる場合がある。なお、このような場合において、例えば一つ一つのピークについて電気泳動時の温度や電圧の変化に関する補正係数を乗じるなどの補正処理を行って、ピークの出現位置等を補正した後にこれらに対比させることも考えられるが、得られたピークの全てにそのような処理を行うのは非常に負担が重く、時間のかかる作業である。また、ピークを構成するフラグメント配列の塩基数とは異なり、本質的に、基準となる絶対的なピークサイズを想定することはできない。

20

【0011】

本発明は前記問題に鑑みてなされたものであり、作成された測定プロファイルデータと参照プロファイルデータとの対比や、作成された測定プロファイルデータ同士の対比を容易に行うことができ、かつ遺伝子の発現状態の解析を容易に行うことができる遺伝子発現解析方法、遺伝子発現解析装置、および遺伝子発現解析プログラムを提供することを課題とする。

30

【課題を解決するための手段】

【0012】

(1) 前記課題を解決した本発明に係る遺伝子発現解析方法は、発現している複数の遺伝子転写産物に由来する複数のピークを一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析方法であって、参照プロファイルデータ取得工程と、測定プロファイルデータ作成工程と、ピーク対応付け工程と、遺伝子発現解析工程を含むことを特徴としている。

【0013】

40

つまり、本発明に係る遺伝子発現解析方法は、参照プロファイルデータ取得工程で、複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られるDNAフラグメントの塩基数相当値の位置と、その位置における前記遺伝子転写産物の転写量相当の検出量に基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ前記遺伝子転写産物の転写産物種情報として、前記第一の波形中の所定のピークと、そのピークが由来する転写産物種と、を同定して記憶した参照プロファイルデータを予め取得しておき、次いで、測定プロファイルデータ作成工程で、複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られる測定対象物となるDNAフラグメントの塩基数相当値の位置と、その位置における前記測定対象物の転写量相当の検出量に基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータを作成する。次いで、ピ

50

ーク対応付け工程で、前記第一の波形および前記第二の波形のうちの少なくとも一方の一部または全部の領域を補正しピーク的位置を調整する補正処理を行い、前記第二の波形中の着目する領域を含む複数のピークと、前記第一の波形における複数のピークと、を対応付けることで、前記着目する領域を含む複数のピークと、当該複数のピークに相当する第一の波形中の複数のピークと、を対応付け、次いで、遺伝子発現解析工程で、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子転写産物の情報を前記転写産物種情報から読み取り、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解析する。

そして、本発明における前記ピーク対応付け工程は、前記参照プロファイルデータ取得工程で取得された前記参照プロファイルデータのピーク的位置と、前記測定プロファイルデータ作成工程で作成された前記測定プロファイルデータのピーク的位置が一致する場合は、一致する前記参照プロファイルデータのピークと前記測定プロファイルデータのピークを対応付け、前記参照プロファイルデータ取得工程で取得された前記参照プロファイルデータのピーク的位置と前記測定プロファイルデータ作成工程で作成された前記測定プロファイルデータのピーク的位置が一部または全部ずれている場合は、これらのうちの少なくとも一方の波形について、これらの波形の類似度が最も高くなるように一部または全部の領域を補正しピーク的位置を調整する補正処理を行った上で、前記参照プロファイルデータのピークと前記測定プロファイルデータのピークを対応付ける。

【0014】

したがって、本発明に係る遺伝子発現解析方法は、取得された参照プロファイルデータと、作成された測定プロファイルデータを波形として表して対比させるため、参照プロファイルデータのピークと、作成された測定プロファイルデータのピークの対応付けを容易に行うことができる。そして、対応付けができた測定プロファイルデータのうちの着目するピークについてその由来となる遺伝子を特定することにより、適切に遺伝子の発現状態を解析することができる。

また、参照プロファイルデータのピーク的位置と測定プロファイルデータのピーク的位置が一致する場合は、これらのピーク的位置を補正処理せずにそのまま対応付けし、これらのピーク的位置がずれている場合のみ、そのピーク的位置を調整する補正処理を行って対応付けをするため、迅速な対応付けが可能となる。

【0015】

(2) 本発明における前記参照プロファイルデータ取得工程は、前記参照プロファイルデータを、既知のプロファイルデータを保存しているデータベースから取得するか、前記転写産物種情報から人工的に作成して取得するか、既知のプロファイルデータ若しくは前記測定プロファイルデータに1つ以上のピークを追加或いは削除することによって取得するか、前記参照プロファイルデータを複数用いて合成することによって取得するか、または、前記測定プロファイルデータを複数用いて合成することによって取得するのが好ましい。

このようにすれば、迅速かつ簡便に参照プロファイルデータを取得して、測定プロファイルデータと対比させることが可能となる。

【0018】

(3) 本発明においては、前記ピーク対応付け工程における補正処理が、ガウス関数に基づく関数近似によって行われるのが好ましい。

このようにすれば、補正処理にかかる負担を軽減しつつ高精度かつ簡便に補正処理を行うことができる。

【0019】

(4) 本発明においては、前記ピーク対応付け工程における補正処理が、前記参照プロファイルデータのピーク的位置を基準として前記測定プロファイルデータのピーク的位置を移動させるか、前記測定プロファイルデータのピーク的位置を基準として前記参照プロファイルデータのピーク的位置を移動させるか、または、前記参照プロファイルデータのピーク的位置と前記測定プロファイルデータのピーク的位置の双方を移動させて、前記測定

10

20

30

40

50

プロファイルデータのピークと前記参照プロファイルデータのピークとを対応付けるのが好ましい。

このようにすれば、参照プロファイルデータピークと測定プロファイルデータのピークを適切に対応付けることができる。

【0020】

(5) 本発明における前記遺伝子発現解析工程は、前記参照プロファイルデータのピークと前記測定プロファイルデータのピークの対応付けができたピークと、対応付けができなかったピークと、が区別できるように表示するとともに、前記参照プロファイルデータのピークに、当該ピークの由来となる遺伝子に関する遺伝子情報が付加されている場合は、当該遺伝子情報を引用することにより前記測定プロファイルデータにおいて対応付けされたピークの遺伝子を特定し、遺伝子の発現状態を解析するのが好ましい。

10

【0021】

このようにすれば、得られた解析結果は、対応付けができなかったピークを容易に見つけることができるとともに、対応付けができた測定プロファイルデータのピークは、その由来となる遺伝子が特定されているので、例えば、後にこれを参照したときに当該ピークが何の遺伝子に由来するか、その塩基配列を解析し直す必要がなくなるなど、ユーザーにとって利用価値の高いものとなる。

【0022】

(6) 本発明における前記遺伝子発現解析工程には、前記ピーク対応付け工程で対応付けができなかったピークについて、当該ピークに関する関連情報を付加する工程が含まれているのが好ましく、(7) 前記関連情報が、前記波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報のうち少なくとも1つを含んでいるのが好ましい。

20

【0023】

このようにすれば、前記したピーク対応付け工程で対応付けができなかったピークについて、当該ピークに関する関連情報を遺伝子発現解析工程で付加することができるので、例えば、後にこれを参照したときに当該波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報などの関連情報を得ることができる。したがって、このようにして得られた解析結果は、ユーザーにとってより利用価値の高いものとなる。

30

【0024】

(8) 本発明に係る遺伝子発現解析装置は、発現している複数の遺伝子転写産物に由来する複数のピークを有する一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析装置であって、複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られるDNAフラグメントの塩基数相当値の位置と、その位置における前記遺伝子転写産物の転写量相当の検出量と、に基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ前記遺伝子転写産物の転写産物種情報として、前記第一の波形中の所定のピークと、そのピークが由来する転写産物種と、を同定して記憶した参照プロファイルデータを予め取得しておく参照プロファイルデータ取得手段と、複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られる測定対象物となるDNAフラグメントの塩基数相当値の位置と、その位置における前記測定対象物の転写量相当の検出量と、に基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータを作成する測定プロファイルデータ作成手段と、前記第一の波形および前記第二の波形のうち少なくとも一方の一部または全部の領域を補正しピーク的位置を調整する補正処理を行い、前記第二の波形中の着目する領域を含む複数のピークと、前記第一の波形における複数のピークと、を対応付けることで、前記着目する領域を含む複数のピークと、当該複数のピークに相当する第一の波形中の複数のピークと、を対応付けるピーク対応付け手段と、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子転写産物の情報を前記転写産物種情報から読み取り、対応付けされた前記測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解

40

50

析する遺伝子発現解析手段を有し、前記ピーク対応付け手段は、前記参照プロファイルデータ取得手段で取得された前記参照プロファイルデータのピークの位置と、前記測定プロファイルデータ作成手段で作成された前記測定プロファイルデータのピークの位置と、が一致する場合は、一致する前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付け、前記参照プロファイルデータ取得手段で取得された前記参照プロファイルデータのピークの位置と、前記測定プロファイルデータ作成手段で作成された前記測定プロファイルデータのピークの位置と、が一部または全部ずれている場合は、これらのうちの少なくとも一方の波形について、これらの波形の類似度が最も高くなるように一部または全部の領域を補正しピークの位置を調整する補正処理を行った上で、前記参照プロファイルデータのピークと、前記測定プロファイルデータのピークと、を対応付けることを特徴としている。

10

【0025】

このようにすれば、取得された参照プロファイルデータと作成された測定プロファイルデータを波形として表わすので、これらの対比を容易に行うことができ、さらに参照プロファイルデータのピークと測定プロファイルデータのピークの対応付けと遺伝子の発現状態の解析を容易に行うことができる。

また、このようにすれば、参照プロファイルデータのピークの位置と、測定プロファイルデータのピークの位置が一致する場合は、これらのピークの位置を補正処理せずにそのまま対応付けし、これらのピークの位置がずれている場合のみ、そのピークの位置を調整する補正処理を行って対応付けをするため、迅速な対応付けが可能となる。

20

【0026】

(9) 本発明における前記参照プロファイルデータ取得手段は、前記参照プロファイルデータを、既知のプロファイルデータを保存しているデータベースから取得するか、前記転写産物種情報から人工的に作成して取得するか、既知のプロファイルデータ若しくは前記測定プロファイルデータに1つ以上のピークを追加或いは削除することによって取得するか、前記参照プロファイルデータを複数用いて合成することによって取得するか、または、前記測定プロファイルデータを複数用いて合成することによって取得するが好ましい。

このようにすれば、迅速かつ容易に参照プロファイルデータを取得して、測定プロファイルデータと対比させることが可能となる。

【0029】

30

(10) 本発明においては、前記ピーク対応付け手段における補正処理が、ガウス関数に基づく関数近似によって行われるのが好ましい。

このようにすれば、補正処理にかかる負担を軽減しつつ高精度かつ簡便に補正処理を行うことができる。

【0030】

(11) 本発明においては、前記ピーク対応付け手段における補正処理が、前記参照プロファイルデータのピークの位置を基準として前記測定プロファイルデータのピークの位置を移動させるか、前記測定プロファイルデータのピークの位置を基準として前記参照プロファイルデータのピークの位置を移動させるか、または、前記参照プロファイルデータのピークの位置と前記測定プロファイルデータのピークの位置の双方を移動させて、前記測定プロファイルデータのピークと前記参照プロファイルデータのピークを対応付けるのが好ましい。

40

このようにすれば、参照プロファイルデータピークと測定プロファイルデータのピークを適切に対応付けることができる。

【0031】

(12) 本発明における前記遺伝子発現解析手段は、前記参照プロファイルデータのピークと前記測定プロファイルデータのピークの対応付けができたピークと、対応付けができなかったピークと、が区別できるように表示するとともに、前記参照プロファイルデータのピークに、当該ピークの由来となる遺伝子に関する遺伝子情報が付加されている場合は、当該遺伝子情報を引用することにより前記測定プロファイルデータにおいて対応付けさ

50

れたピークの遺伝子を特定し、遺伝子の発現状態を解析するのが好ましい。

【0032】

このようにすれば、得られた解析結果は、対応付けができなかったピークを容易に見分けることができるとともに、対応付けができた測定プロファイルデータのピークは、その由来となる遺伝子が特定されているので、例えば、後にこれを参照したときに当該ピークが何の遺伝子に由来するか、その塩基配列を解析し直す必要がなくなるなど、ユーザーにとって利用価値の高いものとなる。

【0033】

(13) 本発明における前記遺伝子発現解析手段には、前記ピーク対応付け手段で対応付けができなかったピークについて、当該ピークに関する関連情報を付加する手段が含まれているのが好ましく、(14) 前記関連情報が、前記波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報のうち少なくとも1つを含んでいるのが好ましい。

10

【0034】

このようにすれば、前記したピーク対応付け工程で対応付けができなかったピークについて、当該ピークに関する関連情報を遺伝子発現解析手段で付加することができるので、例えば、後にこれを参照したときに当該波形の類似度に関する相関係数を基にした評価値、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、およびサンプルの細胞情報や実験情報などの関連情報を得ることができる。したがって、このようにして得られた解析結果は、ユーザーにとってより利用価値の高いものとなる。

20

【0035】

(15) 本発明に係る遺伝子発現解析プログラムは、(1) から (7) に記載の遺伝子発現解析方法をコンピュータに実行させることを特徴としている。

【0036】

このように、コンピュータに(1) から (7) に記載の遺伝子発現解析方法を実行させるので、コンピュータによって参照プロファイルデータと測定プロファイルデータの対比を容易に行うことができ、さらにこれらのピークを対応付けて遺伝子の発現状態を解析することが可能となる。

【発明の効果】

【0037】

本発明の遺伝子発現解析方法によれば、参照プロファイルデータと測定プロファイルデータを波形として表して対比させるため、これらの対比を容易に行うことができ、かつ遺伝子の発現状態の解析を容易に行うことができる。

30

【0038】

また、本発明の遺伝子発現解析装置によれば、参照プロファイルデータと測定プロファイルデータを波形として表して対比させるため、これらの対比を容易に行うことができ、かつ遺伝子の発現状態の解析を容易に行うことができる。

【0039】

そして、本発明の遺伝子発現解析プログラムによれば、参照プロファイルデータと測定プロファイルデータを波形として表して対比させ、さらにこれらのピークを対応付けて遺伝子の発現状態の解析を行わせるようにコンピュータを実行させることができる。

40

【発明を実施するための最良の形態】

【0040】

例えば、網羅的かつ高精度な遺伝子発現解析方法であるHiCEP法による、DNAシーケンサーを用いて得られ、複数のピークを有する波形として表された測定プロファイルデータは、波形の1ピークが、そのサンプルにおける特定の遺伝子に由来するmRNAの存在量(正確には1プロファイル中の相対値)を示している。全てのピークの強度を観測することで、同時に数万種類のmRNAの高精度な発現量を測定することができる。しかし、ピークがどのような塩基配列を持つのか、どのような遺伝子由来なのかは、ピークを分取して塩基配列を解読しなければ知ることができない。公知の膨大な遺伝子情報とHi

50

C E P法による実験の結果を繋ぐためには、ピークの由来遺伝子(mRNA名)を知ることが重要である。なお、遺伝子情報とは、その遺伝子から転写される転写産物名(複数ある場合はリスト等であってもよい)や遺伝子の機能、その他の付属する情報をいい、後記する転写産物情報を包含する。

【0041】

本発明は、HiCEP法などによって作成された測定プロファイルデータを波形として表し、遺伝子情報が調べられている波形(すなわち、参照プロファイルデータの波形)を基準として波形の補正処理を行い、そのピークの位置を整列させること、または、測定プロファイルデータに1つ以上のピークを追加或いは削除して波形の補正処理を行い、整列させることで、HiCEP法による測定プロファイルデータのピークを高精度に遺伝子情報 10
に対応付けて、ピークの由来する遺伝子を特定し、遺伝子の発現状態を解析することのできる遺伝子発現解析方法を具現するものである。

【0042】

以下に、適宜図面を参照して本発明に係る遺伝子発現解析方法、遺伝子発現解析装置、および遺伝子発現解析プログラムを実施するための最良の形態について詳細に説明する。

まず、図1を参照して、本発明に係る遺伝子発現解析方法について説明する。なお、図1は、本発明に係る遺伝子発現解析方法の工程の手順を説明するフローチャートである。

【0043】

本発明に係る遺伝子発現解析方法は、発現している複数の遺伝子転写産物に由来する複数のピークを有する一つの波形として表したプロファイルデータを用いて遺伝子の発現状態 20
を解析する遺伝子発現解析方法である。

図1に示すように、本発明に係る遺伝子発現解析方法は、参照プロファイルデータ取得工程S1と、測定プロファイルデータ作成工程S2と、ピーク対応付け工程S3と、遺伝子発現解析工程S4とを含んでなる。

【0044】

ここで、この技術分野における「発現」とは、一般的にはゲノムDNAからmRNAが転写されてタンパク質として翻訳されることをいうが、本発明における「発現」とは、ゲノムDNAからの遺伝子転写産物、代表的にはmRNAが転写されていることをいうものとする。

また、本発明において「プロファイル」とは、ある条件下におけるサンプルの遺伝子の発現パターン、既知および未知の遺伝子の発現の有無、発現される全ての遺伝子の発現量等を含む情報を示すものであり、「プロファイルデータ」とは、そのような情報を複数のピークを有する一つの波形として表したデータをいう。また、「ピーク」とは、前記した波形の一部を構成し、DNAフラグメントの塩基配列の塩基数と検出量に基づいて、例えば略三角形の形状で表されるものをいう。

【0045】

図1に示すように、参照プロファイルデータ取得工程S1は、複数のmRNAなどの遺伝子転写産物を逆転写したcDNAの一部を増幅して得られるPCR増幅産物などのDNAフラグメントの塩基数相当値の位置と、その位置における遺伝子転写産物の転写量相当の検出量に基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ遺 40
伝子転写産物の転写産物種情報として、第一の波形中の所定のピークと、そのピークが由来する転写産物種とを同定して記憶した参照プロファイルデータを予め取得しておく工程である。ここで、転写産物種情報とは、遺伝子転写産物の種類に関する情報をいい、例えば、ピーク位置、ピーク強度(高さ)が測定され、塩基配列が決定され、さらにその配列名やアクセッション番号が付与され、由来遺伝子名が特定されたものをいう。また、報告された刊行物名や著者名、機関名などの情報が含まれていてもよい。

【0046】

なお、本発明で用いられる参照プロファイルデータには、遺伝子情報を格納したデータベースへのリンク情報をもつピークが含まれて構成されているが、参照プロファイルデータの全てのピークについて由来する転写産物種(および/または遺伝子)が決定されてい 50

なくても参照プロファイルデータとして使用することができる。さらに、複数のプロファイルデータを用いたり、注目するプロファイル部分（ピーク部分）のみを抽出して用いたりしてもよい。また、複数の波形を合成して参照プロファイルデータを作成したり、ピークの位置のみを用いて人工的に参照プロファイルデータを作成したりしてもよい。

【0047】

このような参照プロファイルデータは、既知のプロファイルデータを保存しているデータベースから取得するか、転写産物種情報から人工的に作成して取得するか、既知のプロファイルデータ若しくは測定プロファイルデータに1つ以上のピークを追加或いは削除することによって取得するか、参照プロファイルデータを複数用いて合成することによって取得するか、または、測定プロファイルデータを複数用いて合成することによって好適に取得することができる。つまり、図1に示す、後のピーク対応付け工程S3で測定プロファイルデータの対応付けの対象となるプロファイルデータが得られればよい。したがって、予め用意されたプロファイルデータを用いることに限定されず、前記したように、作成された複数の測定プロファイルデータのうちの一つを参照プロファイルデータとして用い、他のものを測定プロファイルデータとすることができる。

10

【0048】

ただし、当該データベースに保存されている既知のプロファイルデータ中の幾つかのピークに由来する遺伝子について、既に塩基配列が解読され、そのピークについての遺伝子情報、つまり、遺伝子転写産物の転写産物種情報が付加されている場合は、既知のプロファイルデータを参照することにより、これらの情報を同時に得ることができる。したがって、転写産物種情報を得ることができる点で、既知のプロファイルデータを参照する方が、測定プロファイルデータに1つ以上のピークを追加或いは削除して取得したものなどを参照プロファイルデータとして用いるよりも好適である。

20

【0049】

用いる参照プロファイルデータは、ある条件下における特定の細胞（サンプル）と同一の生物種から得られたデータを参照するのが好ましい。ストレインや由来組織（細胞）などサンプルの条件や状態が近ければ近いほど測定プロファイルデータ（測定プロファイルデータについては、後記する測定プロファイルデータ作成工程S2で説明する。）と参照プロファイルデータの類似性が高くなるためより効果的である。しかしながら、本発明においては、類似する細胞の波形を参照プロファイルデータにするなど、サンプルの測定プロファイルデータと、参照する細胞の参照プロファイルデータとを比較的条件に近いものを用いる場合であっても十分効果的である（図2参照）。

30

【0050】

なお、図2の(a)～(h)は、同一生物種由来の異なる細胞株Aと細胞株BについてのHiCEP法（HiCEP法についての具体的な説明は、後記する測定プロファイルデータ作成工程S2で行うこととする。）による測定プロファイルデータを比較した図であって、Xアダプターの $N_1 N_2$ とYアダプターの $N_3 N_4$ の組合せを、(a)は細胞Aについて(AA__AA)、(b)は細胞Bについて(AA__AA)、(c)は細胞Aについて(CT__TC)、(d)は細胞Bについて(CT__TC)、(e)は細胞Aについて(AG__CG)、(f)は細胞Bについて(AG__CG)、(g)は細胞Aについて(AG__CC)、(h)は細胞Bについて(AG__CC)として行ったものである。

40

図2の(a)～(h)に示すように、HiCEP法によって作成された測定プロファイルデータは、異なる細胞株であっても、類似した波形であることが分かる。

【0051】

図1に示すように、次に行う測定プロファイルデータ作成工程S2は、複数のmRNAなどの遺伝子転写産物を逆転写したcDNAの一部を増幅して得られる測定対象物となるPCR増幅産物などのDNAフラグメントの塩基数相当値の位置と、その位置における測定対象物の転写量相当の検出量とに基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータを作成する工程である。

【0052】

50

測定プロファイルデータは、外挿補完や解析対象からの除外、ノイズレベルの指定と除外、重なり合うピークの分離などを行うことでより好適に波形として表すことができる。

外挿補完とは、発現量が大きすぎるために、DNAシーケンサーのセンサが飽和してしまい、波形が台形になったり、中央の落ちくぼんだ巨大なピークになったりするので、解析に先立って本来のピークを推定してピークの形状を補完し、完全なピークとして表すことをいう。

【0053】

また、解析対象からの除外とは、キャピラリーDNAシーケンサーを用いたHiCEP法では何らかの結晶が析出するなどした不純物やゴミが混じた場合に、蛍光を励起させるレーザ光の単純な散乱光によってATGCの各色が同時に測定され、複数のプロファイルの同じ位置に同じ強度でピークが検出されるので、そのような異常なピークを解析対象から除外することをいう。なお、このような異常なピークは、ピークの広がり方()やピークの対象性といったピークの形状からも検出することができる。

10

【0054】

ノイズレベルの指定と除外とは、ノイズとみなすべき小さなピークを指定して除外することをいう。つまり、遺伝子の発現量が非常に少ないために、小さく表されたピークが多数得られるため、そのような小さなピークをノイズとして解析対象から除外することをいう。例えば、遺伝子転写産物の個数が5個/細胞以下となる場合は解析対象から除外することを挙げることができる。

20

【0055】

そして、重なり合うピークの分離とは、大きなピークの裾野に小さなピークが埋もれてしまったり、ピークの肩の部分に別のピークが出てしまったりするために、重なり合うピークは単独のピークと比べて波形がゆがんでしまうため、そのようなピークを分離させることをいう。なお、このような重なり合うピークは、関数近似や波形分析により、波形のゆがみから予測することができるので、例えば、当該ピークの有する情報の一つとしてデータベースなどに記憶させておくことができ、必要に応じて、後記する遺伝子発現解析工程S4でそのピークに関する関連情報などとして付加することができる。

【0056】

また、測定プロファイルデータに重なり合うピークがある場合は以下のようにすることも個々のピークを分離して表すことができる。

30

(1) ピークが重なる部分については、重なり合っているピークごとの寄与分を足し合わせて波形を作成してもよい。この方法によれば、原理的には実際に測定される波形と精度よく一致するので好ましい。

(2) ピークの重なりを無視し、重なっているピークの形状が交差する部分で相手の関数形に切り替えて作成してもよい。この方法によれば、ピークの分離を迅速に行うことができ、ピークが見やすくなるものの、実際の測定波形とは異なる形状になってしまうことがある。

【0057】

なお、これらの方法によっても複数のピークが完全にオーバーラップする場合には最も大きいピークしか見ることができないので、そのような場合にはピークを分取し、塩基配列を解読するなどして解析するのがよい。

40

【0058】

ここで、複数の遺伝子転写産物(例えば、mRNA)を逆転写したcDNAの一部を増幅して得られる測定対象物となるDNAフラグメントの塩基数相当値の位置と、その位置における測定対象物の転写量相当の検出量とに基づいて得られる塩基数相当値の測定範囲を第二の波形として表した測定プロファイルデータは、例えば、図3に示すようにすることによって得ることができる。なお、図3は、測定プロファイルデータを得るまでの手順の一例を示した説明図である。

【0059】

図3(a)に示すように、まず、ポリ(A)RNA(mRNA)11を鋳型として5'

50

末端に、例えばビオチンなどのタグ物質 1 2 が付加された一本鎖 c D N A (First strand cDNA) 1 3 を合成する (a 工程)。次いで、同図 (b) に示すように、 a 工程で合成された一本鎖 c D N A 1 3 を鋳型として二本鎖 c D N A (Second strand cDNA) 1 4 を得る (b 工程)。そして、同図 (c) に示すように、 b 工程で得られた二本鎖 c D N A 1 4 を第 1 の制限酵素 X (例えば、 4 塩基認識の制限酵素) で切断し (c 工程)、同図 (d) に示すように、当該タグ物質に高親和性を有する物質を用いて、 c 工程で得られた c D N A フラグメントから当該タグ物質が付加している c D N A フラグメントを回収する (d 工程)。

【 0 0 6 0 】

次いで、同図 (e) に示すように、 d 工程で回収された c D N A フラグメントの第 1 の制限酵素 X による切断部位へ、当該切断部位の配列に相補的な配列および X プライマーに相補的な配列を含む X アダプター 1 6 を 5 ' 末端側に結合させた c D N A フラグメントを得て (e 工程)、同図 (f) に示すように、 e 工程で得た c D N A フラグメントを、当該 X アダプター 1 6 を切断しない第 2 の制限酵素 Y (例えば、 4 塩基認識の制限酵素) で切断する (f 工程)。そして、同図 (g) に示すように、このタグ物質 1 2 に高親和性を有する物質 1 5 を用いて、 f 工程で得られた c D N A フラグメントから当該タグ物質 1 2 が結合している c D N A フラグメントを取り除くことにより、第 2 の制限酵素 Y による切断部位を含む c D N A フラグメントを回収し (g 工程)、同図 (h) に示すように、 g 工程で回収された c D N A フラグメントの第 2 の制限酵素 Y による切断部位へ、当該切断部位の配列に相補的な配列および Y プライマーに相補的な配列を含む Y アダプター 1 7 を結合させた c D N A フラグメント 1 8 を得る (h 工程)。

【 0 0 6 1 】

次いで、同図 (i) に示すように、この X アダプター 1 6 の配列に相補的な配列を含み、その 3 ' 末端に 2 塩基配列である $N_1 N_2$ (N_1 および N_2 は同一又は異なってもよい、アデニン、チミン、グアニンおよびシトシンからなる群より選ばれる塩基である) を含み、 5 ' 末端に蛍光物質 2 0 を含む X プライマー 1 9 と、当該 Y アダプター 1 7 の配列に相補的な配列を含み、その 3 ' 末端に 2 塩基配列である $N_3 N_4$ (N_3 および N_4 は同一又は異なってもよい、アデニン、チミン、グアニンおよびシトシンからなる群より選ばれる塩基である) を含む Y プライマー 2 1 とからなるプライマーセットを用いて、 h 工程で得られた c D N A フラグメント 1 8 を鋳型とした P C R 反応を行う (i 工程)。そして、同図 (j) に示すように、 i 工程で得られた P C R 増幅産物を D N A シーケンサーにかけて当該 P C R 増幅産物の移動距離 (すなわち、塩基数) および蛍光量 (すなわち、検出量) を検出し (j 工程)、かかる検出結果を波形として表すことで測定プロファイルデータを作成することができる。

【 0 0 6 2 】

このようにすれば、ある条件下における特定の細胞についてのポリ (A) R N A (m R N A) の発現パターン (既知であると未知であるとを問わない。) を、 N_1 、 N_2 、 N_3 、 N_4 の 4 つの塩基の組合せ、つまり $4^4 = 256$ 通りに分類 (サブグループ化) して得ることができる。なお、本発明においてはこれに限定されるものではなく、例えば、 N_1 、 N_2 を N_A 、 N_B 、 N_C とし、 N_3 、 N_4 を N_D 、 N_E 、 N_F とする 6 つの塩基の組合せとしてもよい。なお、この場合は、 $4^6 = 4096$ 通りに分類 (サブグループ化) することができる。もちろん、この場合に用いる制限酵素 X、 Y はいずれも 6 塩基認識の制限酵素を用いる必要がある。

このような測定プロファイルデータの作成方法は、例えば、国際公開番号 W O 2 0 0 2 / 0 4 8 3 5 2 号パンフレットや、特開 2 0 0 5 - 6 5 5 4 号公報に記載されている。

【 0 0 6 3 】

図 1 に示すように、次に行うピーク対応付け工程 S 3 は、前記した第一の波形および前記第二の波形のうちの少なくとも一方の一部または全部の領域を補正しピークの位置を調整する補正処理を行い、前記した第二の波形中の着目する領域を含む複数のピークと、第一の波形における複数のピークとを対応付けることで、着目する領域を含む複数のピーク

と、当該複数のピークに相当する第一の波形中の複数のピークとを対応付ける工程である。

つまり、これらのデータに含まれる個々のピークについて一方または双方のデータの一部または全部の領域を補正しピークの位置を調整する補正処理を行い、同じ位置に有るピークと同じ位置に無いピークを検出し、同じ位置に有るピーク同士を関連付ける。

【0064】

ピーク対応付け工程S3は、例えば、(1)参照プロファイルデータ取得工程S1(図1参照)で取得された参照プロファイルデータのピークの位置と、測定プロファイルデータ作成工程S2(図1参照)で作成された測定プロファイルデータのピークの位置が一致する場合は、一致する参照プロファイルデータのピークと測定プロファイルデータのピークを対応付け、(2)参照プロファイルデータ取得工程S1で取得された参照プロファイルデータのピークの位置と、測定プロファイルデータ作成工程S2で作成された測定プロファイルデータのピークの位置が一部または全部ずれている場合は、これらのうちの少なくとも一方の波形について、これらの波形の類似度が最も高くなるように一部または全部の領域を補正しピークの位置を調整する補正処理を行った上で、参照プロファイルデータのピークと、測定プロファイルデータのピークとを対応付けることにより好適に行うことができる。

10

なお、測定プロファイルデータのピークの位置が一部ずれているとは、参照プロファイルデータのピークの位置に対して、複数の測定プロファイルデータの内のいくつかがずれている場合や、ある測定プロファイルのあるサイズ領域(例えば、200bpから350bpの間)がずれている場合などが該当する。

20

【0065】

ピーク対応付け工程S3における補正処理としては、予め計算基準点を用意し、その基準点間にあるもう1つの基準点の左右を泳動サイズ(塩基数相当値)方向に拡大または縮小して波形相互の評価値(相関係数に類するもの)が向上するように補正処理を行うグローバル補正があり、また、波形のピークが僅かにずれている場合にそのピークの前後の評価値(相関係数に類するもの)を最大にするよう個別の補正量を計算して補正処理を行うローカル補正がある。前者のグローバル補正は、サイズマーカーの認識ずれや実験の条件に由来する相対的に大きな測定揺らぎを吸収することができ、後者のローカル補正は、電気泳動時のcDNAフラグメントの立体構造等に由来するものや、何らかの実験条件の差異に由来する相対的に小さな測定揺らぎを吸収することができる。

30

【0066】

したがって、測定プロファイルデータと参照プロファイルデータの波形の状態に応じてグローバル補正およびローカル補正のうち一方の補正処理、或いは両方を適宜に組み合わせた補正処理を行うことによって、前記したように、これらの波形の類似度が最も高くなるように調整することができる。

【0067】

このような補正処理の一例としては、例えば、(a)参照プロファイルデータのピークの位置を基準として測定プロファイルデータのピークの位置を移動させるか、(b)ある測定プロファイルデータのピークの位置を基準として参照プロファイルデータのピークの位置を移動させるか、または、(c)参照プロファイルデータのピークの位置と測定プロファイルデータのピークの位置の双方を移動させることを挙げることができる。このようにすることによって、測定プロファイルデータの波形と参照プロファイルデータの波形の類似度が最も高くなるように調整することができる結果、参照プロファイルデータのピークと測定プロファイルデータのピークを高精度に対応付けすることが可能となる。

40

なお、前記した補正処理は、補正処理する対象となる波形の数が多ほど高精度に補正することができる。つまり、多数決によって、よりもっともらしい方向へ動かすことができる。したがって、補正処理する対象となる波形の数が十分に多い場合には、計算基準点となるピークの位置を指定してもよいし、そのようなピークの位置を指定しなくてもよい。

50

【 0 0 6 8 】

かかる補正処理は、簡便かつ高精度な補正処理を行うことができるため、ガウス関数に基づく関数近似（ガウス関数近似方式）によって行われるのが好ましいが、例えば、ガウス関数近似方式を基本として、近似による波形寄与分を元のデータから逐次減算して関数近似を繰り返す試行減算方式を併用してもよい。このようにすれば、より適切な補正処理を行うことが可能となる。

【 0 0 6 9 】

ガウス関数近似方式と試行減算方式を併用した補正処理は、例えば、以下のようにして行うことができる。

波形として表した測定プロファイルデータのうちから、主ピーク（1回目の近似で、その近似が確からしいと認められるもの）の寄与を全体の波形から減算し、その残りの部分に対して同様に波形近似を行う。以後、予め定めた範囲に収まるか、予め定めた回数を超えるまでこの処理を繰り返す。なお、かかる補正処理は、裾野の領域には使用せず、ピーク両側の立ち上がり部分を使用して近似を行うと、より高精度に近似させることができる。また、この場合において、補正処理した波形に、補正処理前の波形と重なる測定点ほどの程度存在するかを、その補正の確からしさの評価基準とするとよい。

【 0 0 7 0 】

このような補正処理を行うと、最初に、ガウス関数近似方式による補正処理を行って確かなピークだけをリストアップし、その結果を表示してユーザーの経験則に基づく判断基準との比較を自動的に行い、さらに高次の近似ピークが必要と判断された場合には、試行減算方式による補正処理を行ってより評価値が低いピークも取得するように再度補正処理を行う、といった処理を実施することができる。

【 0 0 7 1 】

なお、飽和ピーク（サチレーション）があった場合は、以下のような補正処理を行うとよい。飽和ピークは、測定器のセンサの飽和状態等により先端が潰れたような形状として検出されるピークであるので、例えば、飽和ピークの両端根元部分である「立ち上がり部分」と「立ち下がり部分」とから波形中央部の先端形状を推定し、本来存在するであろう高さのピークをガウス関数等で作り出す一連の処理が挙げられる。このような補正処理を行うか否かについては、使用する装置のダイナミックレンジを考慮して設定された閾値を超えるかどうかで判断させることができる。

【 0 0 7 2 】

参照プロファイルデータの波形と測定プロファイルデータの波形について前記した補正処理を行い、これらの波形の類似度が最も高くなるように一部または全部のピークの位置を調整して、測定プロファイルデータのピークと参照プロファイルデータのピークを対応付けた一例を図4～6を参照して説明する。なお、図4～6は、12サンプルについて2回繰り返してHiCEP法を行った結果、作成された測定プロファイルデータ（計24波形）を示す図である。

【 0 0 7 3 】

図4は、24波形のピークの位置が局所的にずれてしまった測定プロファイルデータの例であって、（a）は補正処理前、（b）は補正処理後の様子を示す図である。

図4（a）では、中央の2ピークがずれていて、ピーク的位置（塩基数相当値）のみでは両ピークの特定が困難である。しかし、同図（b）の下段に示される波形のうち一番下に表された測定プロファイルデータを参照プロファイルデータとして用い、これを基準として他の23波形のピーク的位置を前記した補正処理によって調整すると、同図（b）の上段の表示部に示されるように参照プロファイルデータの中央の2ピークと、測定プロファイルデータの中央の2ピークとを精度よく対応付けることができる。

【 0 0 7 4 】

また、図5は、24波形のピーク的位置が全体的に大きくシフトしてずれてしまった測定プロファイルデータの例（図5では測定プロファイルデータの一部のみを示している。）であって、（a）は補正処理前、（b）は補正処理後の様子を示す図である。

図5(a)では、24波形のうち6波形がピーク的位置(塩基数相当値)方向に大きくシフトしており、ピークが3つあるように見える。しかし、同図(b)に示すように、24波形のうち1つを参照プロファイルデータとして用い、これを基準として先に述べた6波形を含む他の23波形のピーク的位置を前記した補正処理によって調整するとこれらが揃い、ピーク的位置が2つであることが分かり、これらを精度よく対応付けることができる。

【0075】

図6は、24波形のピーク的位置が近接して複合している測定プロファイルデータの例であって、(a)は補正処理前、(b)は補正処理後の様子を示す図である。

図6(a)では、24波形のピークが近接し、複合しているため、これらの測定プロファイルデータのピーク的位置が少しずれるだけでそれぞれのピークの判定が困難となる。しかし、同図(b)の下段に示される波形のうち一番下に表された測定プロファイルデータを参照プロファイルデータとして用い、これを基準として他の23波形のピーク的位置を前記した補正処理によって調整すると、同図(b)の上段の表示部に示されるように全てのピークが特定の位置で揃い、これらを精度よく対応付けることができる。

【0076】

図7は、10波形の測定プロファイルデータを参考にして、任意の位置にピークを有する波形を人工的に作成して参照プロファイルデータとした場合を示す図である。

図7の上段の表示部に示されるように、10波形の測定プロファイルデータを参考にして、任意の位置にピークを有する波形を人工的に作成して参照プロファイルデータとし、これを基準として10波形の測定プロファイルデータのピーク的位置を前記した補正処理によって調整すると、同図の下段の表示部に示すように全てのピークが揃い、これらを精度よく対応付けることができる。なお、人工的に作成したプロファイルは、前記したように、データベースに登録されているピーク的位置の情報を元に作成することができる。その際、異なる実験条件であれば、発現の大きいピーク、小さいピーク、中間のピークというように強度値(ピークの高さ)を任意に設定することができる。このようにして作成したピークは、補正処理に関して有効に機能し得る。

【0077】

ここで、参照プロファイルデータのピークに、当該ピークの由来に関する遺伝子情報が付加されている場合は、参照プロファイルデータのピークと対応付けすることができた測定プロファイルデータのピークに対して、前記した遺伝子情報を参照プロファイルデータから取得し、当該測定プロファイルデータのピークの遺伝子情報とみなすことができる。

【0078】

図1に示すように、次に行う遺伝子発現解析工程S4は、対応付けされた測定プロファイルデータのピークの由来する遺伝子転写産物の情報を転写産物種情報から読み取り、対応付けされた測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解析する工程である。

遺伝子発現解析工程S4は、例えば、参照プロファイルデータのピークと測定プロファイルデータのピークに対応付けができたピークと、対応付けができなかったピークとが区別できるように表示するとともに、参照プロファイルデータのピークに、当該ピークの由来となる遺伝子に関する遺伝子情報が付加されている場合は、当該遺伝子情報を引用することによって測定プロファイルデータにおいて対応付けされたピークの遺伝子を特定することで遺伝子の発現状態を解析することができる。なお、対応付けができたピークと、対応付けができなかったピークとを区別できるように表示するとは、例えば、一方のピークを色違いで表示したり、一方のピークを明滅するように表示したりすることが挙げられる。

【0079】

遺伝子発現解析工程S4で、参照プロファイルデータのピークと、測定プロファイルデータのピークとの間にピーク対応付けできないピークがあった場合には、そのピークに関する関連情報を付加することにより、より好適な解析結果が得られる。かかる関連情報と

10

20

30

40

50

しては、参照プロファイルデータの波形との類似性に関する相関係数を基にした評価値（図8参照）や、ピーク位置、プライマーセット、発現強度、ピーク形状の特徴、サンプルの細胞に関する情報や、実験条件などの実験情報などを挙げることができ、これらのうち少なくとも1つを含んでいるのがよい。なお、図8は、参照プロファイルデータの波形との類似性に関する相関係数を基にした評価値を図示したものである。

【0080】

図8の中段の表示部に示されているように、参照プロファイルデータに無いピークが測定プロファイルデータに有る場合や、参照プロファイルデータに有るピークが測定プロファイルデータに無い場合がある。この場合、参照プロファイルデータとの関係ではその評価値が低いものとなる。そのため、同図の下段の表示部の中央付近に示されているように、かかる測定プロファイルデータのピークに対する評価値は、他のピークに対する評価値よりも若干低くなる。つまり、ピークサイズに関する信頼度が他のピークよりも若干低くなる。

10

【0081】

もちろん、参照プロファイルデータのピークと対応付けすることのできた測定プロファイルデータのピークに、前記した参照プロファイルデータとの類似性に関する相関係数を基にした評価値などの関連情報を付加してもよい。

なお、このように、関連情報として参照プロファイルデータの波形との類似性に関する相関係数を基にした評価値を付加した場合、かかる評価値は同時に、測定プロファイルデータに付加された転写産物種情報の対応付けの確からしさに関する評価を与えることにもなる。

20

【0082】

以上に述べたように、本発明の遺伝子発現解析方法によれば、参照プロファイルデータと測定プロファイルデータを波形として表して対比させるため、これらの対比を容易に行うことができ、かつ遺伝子の発現状態の解析を容易に行うことができる。また、従来は、測定プロファイルデータのピークがどのような遺伝子転写産物（mRNA）由来のものなのか、毎回、ピークを分取して塩基配列を決定し、任意のデータベースによって類似性検索などを行って遺伝子を決定しなければならなかったが、対応付けに用いた参照プロファイルデータのピークに遺伝子情報が付加されている場合は、これにより測定プロファイルデータのピークが由来する遺伝子を同定することができるので、前記した作業を省くことができる。

30

【0083】

次に、図9を参照して、本発明に係る遺伝子発現解析装置について説明する。なお、図9は、本発明に係る遺伝子発現解析装置の構成の一例を示すブロック図である。

【0084】

本発明に係る遺伝子発現解析装置Aは、発現している複数の遺伝子転写産物に由来する複数のピークを有する一つの波形として表したプロファイルデータを用いて遺伝子の発現状態を解析する遺伝子発現解析装置であって、図9に示すように、参照プロファイルデータ取得手段1と、測定プロファイルデータ作成手段2と、ピーク対応付け手段3と、遺伝子発現解析手段4とを有する。

40

なお、本発明に係る遺伝子発現解析装置Aの参照プロファイルデータ取得手段1、測定プロファイルデータ作成手段2、ピーク対応付け手段3および遺伝子発現解析手段4はそれぞれ、図1に示す、本発明に係る遺伝子発現解析方法の参照プロファイルデータ取得工程S1、測定プロファイルデータ作成工程S2、ピーク対応付け工程S3および遺伝子発現解析工程S4に対応するものであるため、以下の説明においては、前記した内容と重複する内容についての説明を省略することとする。

【0085】

かかる遺伝子発現解析装置Aは、例えば、一般的に使用される汎用コンピュータやワークステーションなどを使用することができ、図示しない接続手段によって外部のキャピラリーDNAシーケンサーなどのDNAシーケンサーDSや、インターネットなどの通信ネ

50

ットワークNWを介して参照プロファイルデータベースDB1、遺伝子情報データベースDB2などと接続されており、適時、必要とされる情報を取得することができる。また、遺伝子発現解析装置Aには、ハードディスクドライブなどの記憶手段5が設けられており、必要に応じて前記した各手段でした参照データや作成したデータ、解析結果などを記憶させたり、読み出ししたりすることができる。

【0086】

参照プロファイルデータ取得手段1は、通信ネットワークNWを介して参照プロファイルデータベースDB1から、後記する測定プロファイルデータ作成手段2で作成された測定プロファイルデータの参照対象となる参照プロファイルデータを予め取得する。参照プロファイルデータは、前記したように、複数の遺伝子転写産物を逆転写したcDNAの一部を増幅して得られるDNAフラグメントの塩基数相当値の位置と、その位置における遺伝子転写産物の転写量相当の検出量とに基づいて得られる塩基数相当値の参照範囲を第一の波形として表わし、かつ遺伝子転写産物の転写産物種情報として、第一の波形中の所定のピークと、そのピークが由来する転写産物種とを同定して記憶したデータである。なお、かかる参照プロファイルデータに含まれるピークに、遺伝子名などの遺伝子情報が付加されている場合、または付加するような設定がされている場合は、当該遺伝子情報を、通信ネットワークNWを介して遺伝子情報データベースDB2から取得してもよい。

10

【0087】

ここで、参照プロファイルデータベースDB1や遺伝子情報データベースDB2は一意のデータベースに限定されるものではなく、参照プロファイルデータとして用いることのできるデータが異なる複数のデータベースに格納された、いわゆるデータベース・プールのようなものも含まれる。このような場合、参照プロファイルデータを取得するためには、複数のデータベースから適宜必要なデータを取得することになる。なお、どのようなデータを参照プロファイルデータとするかはユーザーが任意に決定することができる。例えば、最も沢山のピークがあり、最も近い生物材料、または実験条件であり、最も多くの遺伝子情報に対応付けられているものを、必要であれば複数のデータベースから取得して参照プロファイルデータとすることができる。これは例えば、測定プロファイルデータが1つか2つといった少数である場合、1つの参照プロファイルデータのみを取得しただけであると、波形の補正処理の精度が良くないことが有り得る。そのような場合は、前記したように比較的近いデータを複数取り出してきて参照プロファイルデータとし、波形の補正処理を行い、ピーク対応付けに使用するとよい。

20

30

【0088】

参照プロファイルデータベースDB1としては、例えば、独立行政法人放射線医学総合研究所(日本国)の提供するHiCEP用のデータベース(URL: <http://133.63.22.11/peakdb/query?request=dbmain&lang=ja>)などを挙げるることができる。また、遺伝子情報データベースDB2としては、例えば、NCBI(National Center for Biotechnology Information(国立バイオテクノロジー情報センター)(アメリカ合衆国))URL: <http://www.ncbi.nlm.nih.gov/>の提供するGenBank(URL: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>)、一塩基多型(SNP)のデータベースであるdbSNP(URL: <http://www.ncbi.nlm.nih.gov/projects/SNP/>)、EST(Expressed Sequence Tag)のデータベースであるdbEST(URL: <http://www.ncbi.nlm.nih.gov/dbEST/>)、文献データベースであるMEDLINE(PubMed(URL: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed>))などのデータベースを挙げるこ

40

【0089】

そして、測定プロファイルデータ作成手段2は、例えば、図3を参照して説明したように、測定対象となる複数の遺伝子転写産物を逆転写したcDNAの一部をPCR法などにより増幅して得られる測定対象物となるDNAフラグメント(PCR増幅産物)をDNAシーケンサーDSにかけることにより測定された塩基数相当値の位置と、その位置における測定対象物の転写量相当の検出量とを含むシーケンス結果SRを取得することにより、

50

かかるシーケンス結果SRに基づいて得られる塩基数相当値の測定範囲を例えば、横軸を塩基数とし、縦軸を検出量としたグラフの形式で複数のピークを有する第二の波形として表した測定プロファイルデータを作成する。

【0090】

ピーク対応付け手段3は、前記した第一の波形および前記した第二の波形のうちの少なくとも一方の一部または全部の領域を補正しピークの位置を調整する補正処理を行い、第二の波形中の着目する領域を含む複数のピークと、第一の波形における複数のピークとを対応付けることで、前記着目する領域を含む複数のピークと、当該複数のピークに相当する第一の波形中の複数のピークとを対応付ける。

【0091】

遺伝子発現解析手段4は、対応付けされた測定プロファイルデータのピークの由来する遺伝子転写産物の情報を転写産物種情報から読み取り、対応付けされた測定プロファイルデータのピークの由来する遺伝子を特定することで遺伝子の発現状態を解析する。なお、ピーク対応付け手段3で対応付けされた測定プロファイルデータのピークに遺伝子情報を付加する場合は、この遺伝子発現解析手段4で遺伝子情報データベースDB2にアクセスして対応付けされたピークに関する遺伝子情報を付加するようによい。

【0092】

そして、前記のようにして解析された解析結果をディスプレイやプリンターなどの表示手段6に出力することで、ユーザーに遺伝子の発現状態を提示することができる。

【0093】

そして、本発明に係る遺伝子発現解析プログラムは、CD-ROM、フレキシブルディスク等のコンピュータ読み取り可能な記録媒体(図示せず)に記録され、例えば、遺伝子発現解析装置Aと接続された記録媒体駆動装置(図示せず)によって、当該記録媒体から遺伝子発現解析プログラムを読み出して記憶手段5にインストールすることにより、コンピュータに、図1に示す参照プロファイルデータ取得工程S1、測定プロファイルデータ作成工程S2、ピーク対応付け工程S3および遺伝子発現解析工程S4を実行させるようによい。

【0094】

また、遺伝子発現解析プログラムが通信ネットワークNWを介して接続された他のコンピュータ(サーバ)に記憶されている場合、通信ネットワークNWに接続された遺伝子発現解析装置(クライアント)が、当該他のコンピュータから通信ネットワークNWを介して遺伝子発現解析プログラムをダウンロードすることにより、コンピュータに参照プロファイルデータ取得工程S1、測定プロファイルデータ作成工程S2、ピーク対応付け工程S3および遺伝子発現解析工程S4を実行させるようによい。

【0095】

以上、本発明の遺伝子発現解析方法、遺伝子発現解析装置、および遺伝子発現解析プログラムについて、発明を実施するための最良の形態により詳細に説明したが、本発明の趣旨はこれに限定されるものではなく、特許請求の範囲の記載に基づいて広く解釈されなければならないことはいうまでもない。

【図面の簡単な説明】

【0096】

【図1】本発明に係る遺伝子発現解析方法の工程の手順を説明するフローチャートである。

【図2】(a)~(h)は、異なる細胞株Aと細胞株BについてのHiCEP法による測定プロファイルデータを比較した図である。

【図3】測定プロファイルデータを得るまでの手順の一例を示した説明図である。

【図4】24波形のピークの位置が局所的にずれてしまった測定プロファイルデータの例であって、(a)は補正処理前、(b)は補正処理後の様子示す図である。

【図5】24波形のピークの位置が全体的に大きくシフトしてずれてしまった測定プロファイルデータの例であって、(a)は補正処理前、(b)は補正処理後の様子示す図であ

10

20

30

40

50

る。

【図6】24波形のピークの位置が近接して複合している測定プロファイルデータの例であって、(a)は補正処理前、(b)は補正処理後の様子を示す図である。

【図7】10波形の測定プロファイルデータを参考にして、任意の位置にピークを有する波形を人工的に作成して参照プロファイルデータとした場合を示す図である。

【図8】参照プロファイルデータの波形との類似性に関する相関係数を基にした評価値を図示したものである。

【図9】本発明に係る遺伝子発現解析装置の構成の一例を示すブロック図である。

【符号の説明】

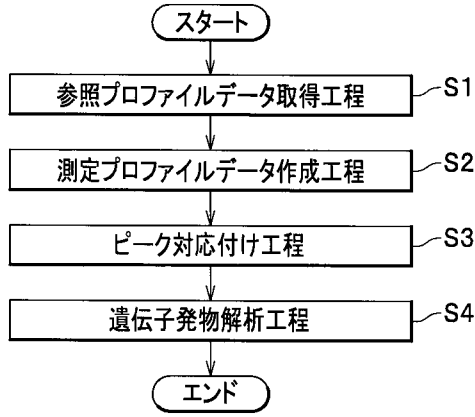
【0097】

S 1	参照プロファイルデータ取得工程	
S 2	測定プロファイルデータ作成工程	
S 3	ピーク対応付け工程	
S 4	遺伝子発現解析工程	
A	遺伝子発現解析装置	
1	参照プロファイルデータ取得	
2	測定プロファイルデータ作成	
3	ピーク対応付け工程	
4	遺伝子発現解析工程	
5	記憶手段	20
6	表示手段	
D S	DNAシーケンサー	
S R	シーケンス結果(DNAフラグメント(PCR増幅産物)の塩基数と検出量)	
N W	通信ネットワーク	
D B 1	参照プロファイルデータベース	
D B 2	遺伝子情報データベース	

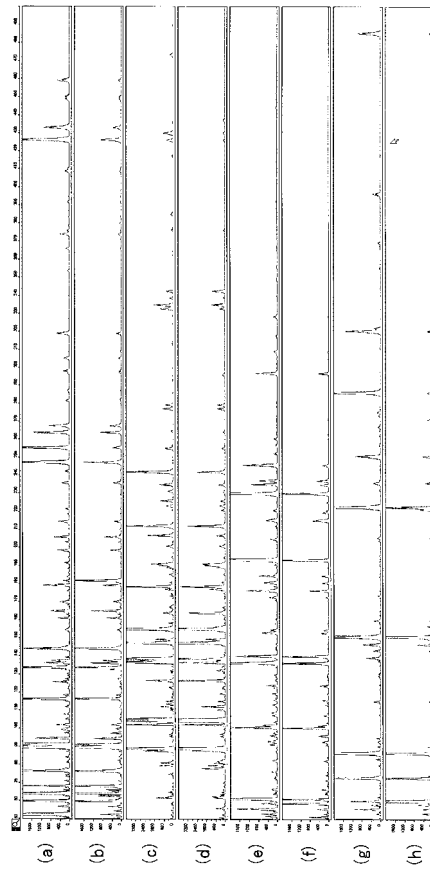
10

20

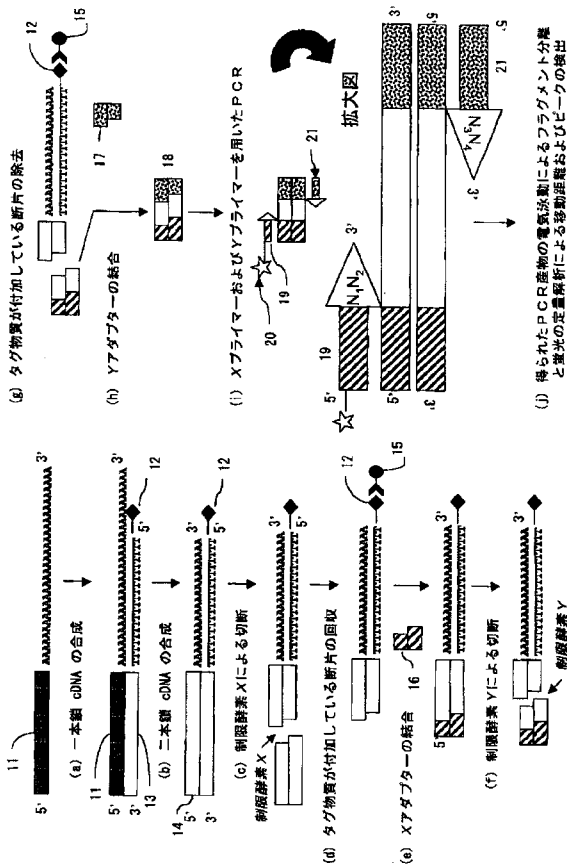
【図 1】



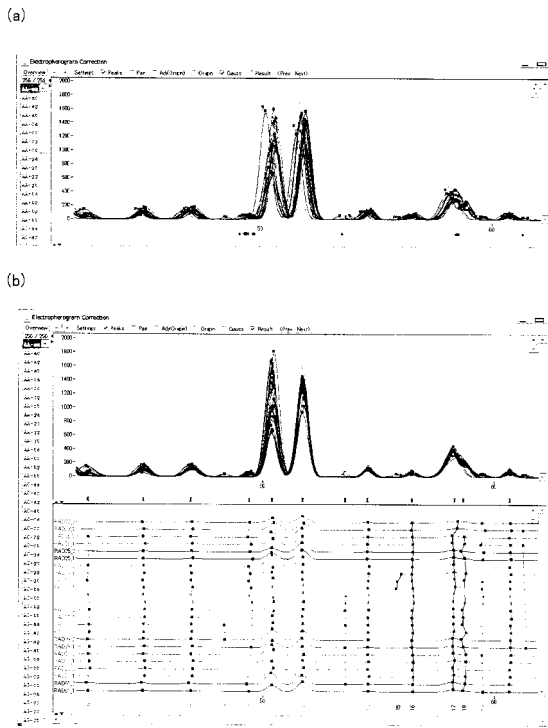
【図 2】



【図 3】

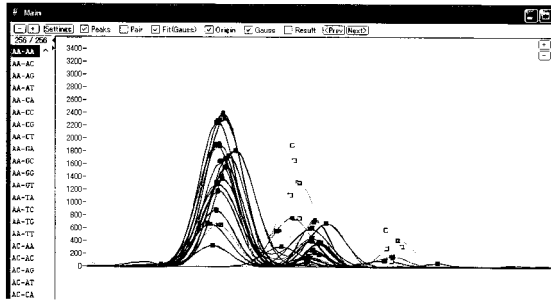


【図 4】

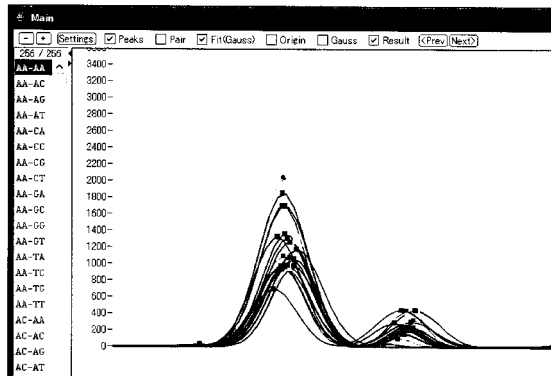


【 5 】

(a)

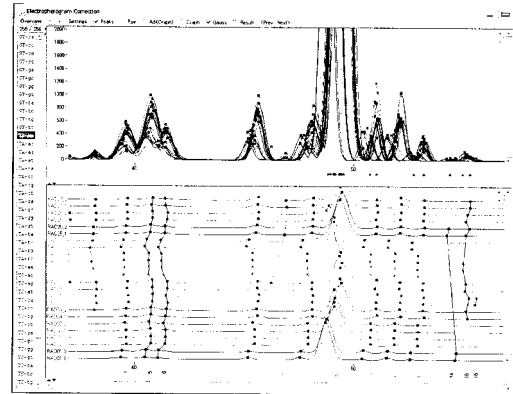


(b)

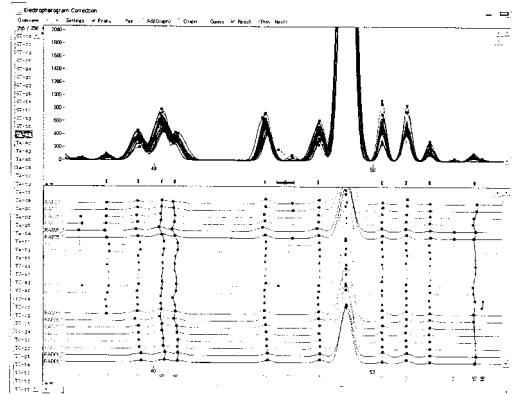


【 6 】

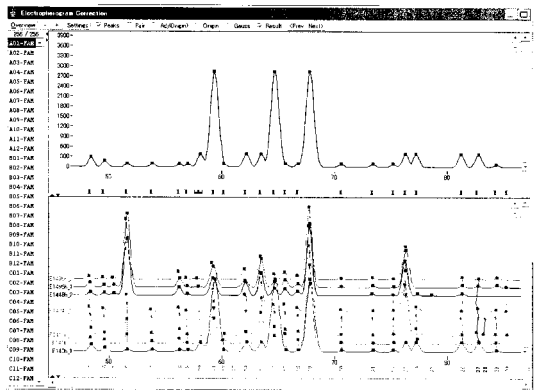
(a)



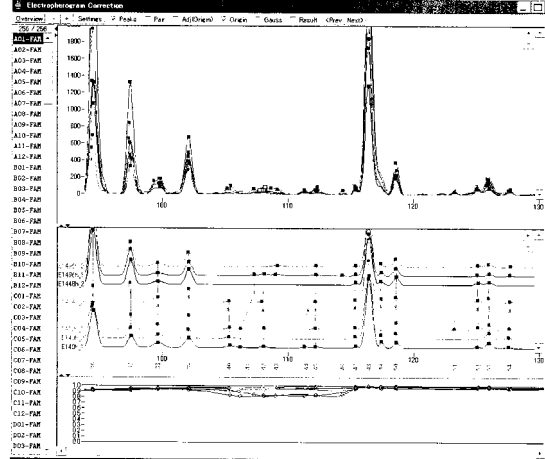
(b)



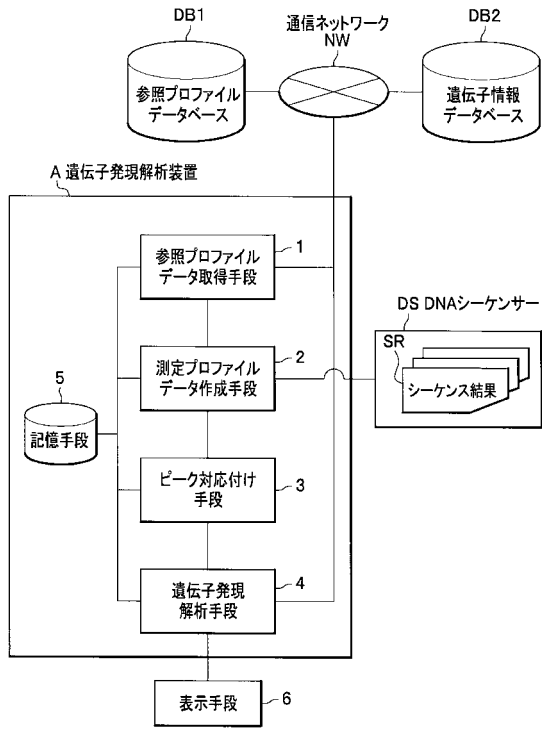
【 7 】



【 8 】



【図9】



フロントページの続き

- (56)参考文献 特開2005-250615(JP,A)
特開2008-226095(JP,A)
特表2000-500647(JP,A)
特開2002-325586(JP,A)
国際公開第02/048352(WO,A1)
特開2002-335963(JP,A)
Kadota, K., A normalization strategy applied to HiCEP (an AFLP-based expression profiling) analysis: Toward the strict alignment of valid fragments across electrophoretic patterns, BMC Bioinformatics, 2005年 3月 6日, Vol.6, p.43
Kadota, K., GOGOT: a method for the identification of differentially expressed fragments from cDNA-AFLP data, Algorithms for Molecular Biology, 2007年 3月30日, Vol. 2, p.5
安倍真澄, 高精度遺伝子発現プロファイル解析(HiCEP)技術とその可能性, 放射線科学, 実業広報社, 2004年10月25日, 第47巻, 第10号, p.26-35

(58)調査した分野(Int.Cl., DB名)

G06F 19/10

JSTPlus/JMEDPlus/JST7580(JDreamIII)

PubMed