

Title	トランスクリプトーム解析のための遺伝子発現プロファイルの分類手法および類似構造探索に関する研究
Author(s)	瀬尾, 茂人
Citation	大阪大学, 2006, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/46648
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	瀬尾 茂人
博士の専攻分野の名称	博士 (情報科学)
学位記番号	第 20508 号
学位授与年月日	平成 18 年 3 月 24 日
学位授与の要件	学位規則第 4 条第 1 項該当 情報科学研究科バイオ情報工学専攻
学位論文名	トランスクリプトーム解析のための遺伝子発現プロファイルの分類手法 および類似構造探索に関する研究
論文審査委員	(主査) 教授 松田 秀雄 (副査) 教授 柏原 敏伸 教授 赤澤 堅造 教授 清水 浩

論文内容の要旨

近年、技術革新や、公共データベースの充実などの理由により、ゲノムワイドな遺伝子発現に関する大量実験情報が比較的手軽に得られる時代になってきており、遺伝子の多様なメカニズムを解明するため、これらを用いたトランスクリプトーム解析の重要性が高まっている。トランスクリプトーム解析の基本となるのは、発現プロファイルであり、これは遺伝子の発現量を様々な条件下において測定した結果得られる集積データのことである。つまり発現プロファイルは、いつでもどこでものくらいそれぞれの遺伝子が発現しているかを調べたデータであると言える。網羅的な発現プロファイルの測定には、塩基配列決定技術 (シーケンシング) による EST (Expression Sequence Tag) や CAGE (Cap Analysis Gene Expression)、ハイブリダイゼーションを利用するマイクロアレイによるアプローチが広く利用されている。

本研究では、マイクロアレイにより測定された発現プロファイルの解析によく利用されるクラスタリングの問題点に着目し、それを解決するための手法の提案、また、新しい方向性のトランスクリプトーム解析として、CAGE タグを用いたゲノム上での発現構造の類似性探索手法の提案を行った。

クラスタリングとは、多数のデータを、類似した特徴量を持つデータごとにまとめることにより、いくつかのグループに分類する手法である。クラスタリングによって得られる分類された各グループをクラスタと呼ぶ。発現プロファイルを基に遺伝子のクラスタリングを行うことによって、同じ時期や組織で発現する遺伝子のクラスタが得られる。クラスタリングは、機能未知の遺伝子の機能を、それと共発現している既知の機能を持つ遺伝子を手がかりに機能推定したり、特定の条件下で特徴的に発現している遺伝子群を見つけ解析したりするときによく用いられる。しかし、クラスタリングの中でも特によく使用される平均連鎖法では、発現プロファイル間の類似度のばらつきが大きいときでも強制的に平均をとってクラスタとしてまとめてしまうという欠点がある。この問題に対処するために、 p -準完全グラフと呼ばれるグラフ構造に基づく発現プロファイルのクラスタリング手法を提案した。提案手法を実際の発現プロファイルに対して適用し、従来手法との比較、考察を行った結果、本手法は従来手法よりも高い選択度を得ることができており、従来の方法ではうまく分類できないようなデータに対しても有効であると考えられる。また、マイクロアレイによって測定された発現プロファイルには、様々なノイズやバイアスが影響を与えることを考慮し、相対エントロピーや局所的な類似性を評価するための尺度をクラスタリングに導入した。この結果として、 p -準完全グラフ

構造に基づくクラスタリング手法の精度をより向上させることができている。

さらにトランスクリプトーム解析の新しいアプローチとして、CAGE を用いたゲノム上の発現構造の類似性探索手法の提案を行った。CAGE とは発現量と正確な転写開始点が得られる、塩基配列決定技術を利用した測定手法である。本研究では、この発現量をゲノムにマッピングし、さらに正規化・離散化して文字列へと変換することで、従来の文字列比較のように動的計画法を用いて、ゲノム上の発現パターンの類似性探索を行うことができる手法を提案した。これは、似たメカニズムを持つゲノム上の領域は似た発現のパターンを示す可能性があり、各遺伝子の発現量だけではなく、ゲノム上の位置情報を考慮に入れることは、複雑な転写制御・発現調節のネットワークのメカニズムを解明する上で重要であると考えられる。

論文審査の結果の要旨

本論文では、発現プロファイルの解析によく利用されるクラスタリングの問題点に着目し、それを解決するための手法の提案、また新しい方向性のトランスクリプトーム解析として CAGE (Cap Analysis Gene Expression) タグを用いたゲノム上での発現構造の類似性探索手法の提案を行っている。

まず、よく用いられているクラスタリング手法である平均連鎖法について、発現プロファイル間の類似度のばらつきが大きいときでも強制的に平均をとってクラスタとしてまとめてしまうという問題点を指摘し、この問題に対処するためにグラフ構造を利用したクラスタリング手法を提案している。そして、提案手法を実際の発現プロファイルに対して適用した結果、従来手法よりも誤って分類する割合が小さいという結果が示されている。各種の癌を特徴付ける遺伝子群をクラスタとして抽出することもできていることから、提案手法の有効性が示されている。また、マイクロアレイによって測定された発現プロファイルには、様々なノイズやバイアスが影響を与えることを考慮し、相対エントロピーや局所的な類似性を評価するための尺度をクラスタリングに導入している。この結果として、グラフ構造に基づくクラスタリング手法の精度をより向上させることができたことも報告されている。さらにトランスクリプトーム解析の新しいアプローチとして、CAGE を用いたゲノム上の発現構造の類似性探索手法を提案している。CAGE (Cap Analysis Gene Expression) とは発現量と正確な転写開始点が得られる、塩基配列決定技術を利用した測定手法である。提案手法は、CAGE タグをゲノムにマッピングして文字列へと変換することで、発現量と発現位置を同時に比較することを可能にしている。似たメカニズムを持つゲノム上の領域は似た発現のパターンを示す可能性があり、各遺伝子の発現量だけではなく、ゲノム上の位置情報を考慮に入れることは、複雑な転写制御・発現調節のネットワークのメカニズムを解明する上で重要である。提案手法をマウスの CAGE ライブラリに対して適用することにより、各組織における各染色体の発現構造の網羅的な類似性探索を行っている。その結果として強く発現している遺伝子クラスタを含む領域間に類似性を検出したことなどが報告され、有効性が示されている。

以上により、本論文の成果は計算機による遺伝子機能解析に関する研究の発展に貢献するものである。よって博士(情報科学)の学位論文として価値あるものとして認める。