

氏名	やまにしよしひろ 山 西 芳 裕
学位(専攻分野)	博 士 (理 学)
学位記番号	理 博 第 2933 号
学位授与の日付	平成 17 年 3 月 23 日
学位授与の要件	学位規則第 4 条第 1 項該当
研究科・専攻	理学研究科 生物科学専攻
学位論文題目	異質なゲノムデータの相関解析法の開発とタンパク質ネットワーク予測への応用
論文調査委員	(主査) 教授 金久 實 教授 森 和俊 教授 七田 芳則

論 文 内 容 の 要 旨

ゲノム研究およびプロテオーム研究の進展とともに、遺伝子やタンパク質に関する、配列情報、発現情報、進化情報、局在情報、相互作用情報などがゲノムワイドに得られるようになった。ゲノム情報から生命システムへの情報構築原理を明らかにするためには、これらの個々のデータ解析はもちろん、データ間の比較や統合を行い、新しい生物学的な知見を得るための統合的解析が必要である。本研究では、これらの異質なゲノムデータを統一的に扱い、様々な解析をするための方法論の開発を行った。

まず、ゲノムデータの一つである系統プロファイルの解析を行った。生物種間の比較という観点から見ると、各生物種のプロファイルは、進化の過程において遺伝子の得失を表す情報と解釈できる。本研究では、遺伝子の得失パターンに基づく生物種の分類の結果、9つの主要な生物種グループを抽出することができた。そして、各生物種グループに特徴的な遺伝子群を同定し、生物学的機能との関係を検証した。

次に、異質なゲノムデータの相関解析法を開発した。同じ遺伝子またはタンパク質に関して、配列情報や、遺伝子発現、パスウェイ(タンパク質ネットワーク)などのデータが得られたときに、これらの異なる生物学的属性間の相関を解析することが重要である。本研究では、カーネル正準相関分析という手法を拡張することによって、複数の異質なゲノムデータ間の相関を解析し、その相関に寄与する遺伝子群を抽出する手法を開発した。実際に、提案した手法を大腸菌のオペロン構造の検出法として適用した。バクテリアのオペロン構造は、パスウェイ、ゲノム上の並び、遺伝子発現という3つの異質なデータ間の生物学的相関の一例であると考えられることができる。そこで、代謝パスウェイ上の機能的な遺伝子産物間の関係、染色体上での遺伝子の隣接関係、マイクロアレイ実験で共発現する遺伝子間の関係を表す3つのデータを用いて、オペロンに属していると考えられる遺伝子群を探索的に抽出した。大腸菌のオペロンデータベースとの比較の結果、抽出した遺伝子群は、既知のオペロンに属する遺伝子群とよく対応していることが確認できた。

最後に、開発したカーネル正準相関分析のモデルを応用し、タンパク質間の機能的なネットワークの予測を行った。ネットワーク情報が既知のタンパク質セットで、ゲノムデータとネットワークの相関モデルを構築し、そのモデルを当てはめることによって、ネットワーク情報が未知のタンパク質間のネットワークを予測する方法を提案した。まず、KEGG データベースに登録されている出芽酵母のタンパク質ネットワークを正解データとみなし、それを出芽酵母のマイクロアレイ遺伝子発現データ、酵母2ハイブリッドによるタンパク質相互作用情報、タンパク質の細胞内局在情報、系統プロファイルの4つのデータを用いて予測し、正解ネットワークの復元度合いで提案手法の予測精度を検証した。次に、出芽酵母の網羅的なタンパク質間ネットワークを推定し、代謝パスウェイに存在する missing 酵素遺伝子の推定や、機能未知のタンパク質の機能推定などを行った。また、もう一つの適用例として、緑膿菌のタンパク質ネットワークの予測を行い、リジン代謝パスウェイの再構築を行った。ここでは、染色体上での遺伝子間の近さ、系統プロファイルによるタンパク質間の進化的な類似度を用いて、タンパク質の機能ネットワークを推定し、リジン分解系のパスウェイ上にあると思われる酵素遺伝子を予測した。

予測した候補遺伝子群のいくつかについて、実際に実験で酵素活性を示し、予測結果の妥当性を示唆した。

論文審査の結果の要旨

ポストゲノム時代と呼ばれる今日では、プロジェクトの重点は、従来のゲノムプロジェクトが行ってきたような生体関連分子のカタログ化の作業から、生命現象を一つの情報システムとして解析する研究へ移行しつつある。このような背景の下、ポストゲノム解析によって生み出される大量の実験データの解析、そのデータから推定される巨大な分子間相互作用の解析などを、情報科学的手法を用いて行うことが必要不可欠となっている。それら大量データから個々のタンパク質の機能を予測するための方法論に関する研究は今日まで盛んに行われているが、本論文では、タンパク質間の機能的関係の予測法に焦点を当てている。ここでは、論文の内容を、異質なゲノムデータ間の相関関係を解析する手法の開発を行った部分と、その手法を応用してタンパク質間のネットワーク予測法を提案した部分とに分けて評価した。

申請者が開発した相関解析法は、カーネル正準相関分析という統計解析法を基本にしており、このモデルをいくつか拡張することによって、同じ遺伝子あるいはタンパク質の情報を表すデータ間の相関解析を可能にしている。申請者が拡張した点は、データ構造の異なるゲノム情報を統一的に扱うための枠組み、複数のデータの統合という2点である。グラフ、文字列、数値ベクトルなどのデータの異質性は、異なる生物学的属性間関係を解析するときの問題点で、これまで標準的な方法は無かった。申請者の方法によって、どんな構造のデータも類似度行列として統一的に扱え、異質なデータを容易に統合できるようになったことは評価に値する。実際に、代謝パスウェイ、ゲノム上での並び、遺伝子発現データに基づくオペロン抽出において、その有効性が示されており一定の成果をあげていると認められる。

タンパク質ネットワークの予測法としての独自性は、相関解析で用いたカーネル正準相関分析のモデルを応用し、教師付き学習の枠組みにおいてネットワーク推定問題を考えた点にある。教師付きの枠組みでのネットワーク予測はこれまでに例がなく、全く新しい方法である。実際に性能面でも、先行研究の方法論よりも大幅な予測精度の向上がみられ、信頼性の高いタンパク質ネットワーク予測法を開発したことは、タンパク質間における未知の機能的関係の発見につながるため意義は大きい。代謝パスウェイでは、経路はあると分かっているが、各触媒反応に関わる酵素の遺伝子が同定されていないというmissing酵素が多数ある。これまでは、配列類似性に基づいて酵素遺伝子を同定することが一般的であり、他の生物種で酵素遺伝子が同定されていない酵素の場合、対象の酵素遺伝子を探す方法は無かった。申請者の手法では、配列類似性を必要とせずに、遺伝子の共発現や進化情報、タンパク質の局在情報などの網羅的な実験データを使って、missing酵素の遺伝子候補の推定ができるという点が特長である。Missing酵素は、既知の代謝パスウェイに多数存在するので、提案手法で網羅的にmissing酵素遺伝子を同定することで、新しい生物学的発見を得ることが期待でき、実用性という面からも評価できる。

以上申請者が行った、異質なゲノムデータの相関解析法の開発およびタンパク質の機能ネットワーク予測は、様々なゲノム情報や実験データを用いた全く新しい研究手法を導くものであり、基礎的な研究成果としても実用的な側面においても高く評価できる。

よって、本論文は博士(理学)の学位論文として価値あるものと認める。なお、論文内容とそれに関連した試問の結果、合格と認めた。