

論文の内容の要旨

論文題目 知識情報を活用した時系列解析の方法論の開発と遺伝子発現データへの応用

氏名 北上 純一

生物学の課題の一つは生体を構成する遺伝子、タンパク質、分子がどう相互作用し、細胞機能のどのように関与しているかを理解することである。遺伝子やタンパク質は多様なレベルで相互作用をしているため、単に個々の機能を解析するだけではなく、それらの中での相互作用を明らかにして、システム全体として複雑な作用を理解することが重要である。遺伝子全体(ゲノムワイド)な遺伝子の発現情報を得る手段として、マイクロアレイを用いることが一般的に行われるようになってきた。マイクロアレイによる計測技術が進歩するにつれて、時間および空間分解能の高い大量の発現データを効率的に解析する方法論の開発が望まれている。

マイクロアレイ技術は何万ものゲノム全体の遺伝子発現データを効率良くモニタリングするハイスループットな技術として、1990年後半から急速に利用されるようになってきた。最近では、この技術を用いて遺伝子発現データから特異的な遺伝子集合を抽出し、特定の疾患で特異的な発現を示す遺伝子集合(バイオマーカー)や、特定の生物学プロセスと関連のある遺伝子集合などと比較を行い、実験対象の生体内での現象の解明、創薬に用いられている。

マイクロアレイは容易にゲノムワイドな情報を取得できる利点を持つ反面、ノイズが大きくなることがあるという欠点をもつ。マイクロアレイのノイズには実験の過程で生じるノイズと生物学的ノイズがあると考えられている。特に、後者の生物学的ノイズは個々の遺伝子の転写までのメカニズムにそのノイズのレベルが依存するので、統計的な手法などを用いて修正、評価することが困難である。これらノイズが発現遺伝子から特異的な発現を示す集合を抽出する際に、大きな影響を及ぼすためにノイズに強い解析が遺伝子解析において課題となっている。

時間系列で遺伝子発現サンプルを取得したデータに対しては、各遺伝子が示す時間軸方向のデータパターンが類似したものをグループ(クラスター)とするクラスタリング技法により、特異的な発現遺伝子集合を抽出するのが一般的である。クラスタリングにおいてもやはりノイズの影響を強く受けてしまうために正確なクラスターを抽出できない課題がある。最近では、ノイズに強いいくつかの手法が提案されているが、マイクロアレイのノイズが統計的に処理しにくい性

質があるために大きな改善は難しい。

クラスタリングでのノイズによる誤差を解決する他のアプローチは既にある他の生物学的知識情報をアルゴリズムに取り込み不足する情報を補うことである。これによって発現データのみでは曖昧になっていたクラスタリングを補正しくモジュールとして解析することが期待できる。生物学分野の知識情報のデータベースの中で最も情報を多く含むデータベースはGene Ontology (GO)である。GOは現在において急速に充実してきている情報を豊富に含んだ知識ソースで、生物学の機能やプロセス用語を体系的に蓄積している。GOの用語は遺伝子と関連付けられるためこの関係性を用いることによって知識情報を遺伝子に付加することが可能になる。GOは遺伝子の機能と密接な関係を持つことが知られているので、2遺伝子間の類似度をクラスタリング解析時に遺伝子間情報として組み込む方法を用いた。

遺伝子間の付加情報をうまく取り込むためにクラスタリング技術としては遺伝子発現のネットワークを作成してクラスタリングを行う方法を用いた。パターンの相関性が高い関係性を(エッジ)枝としてネットワークを構築し、ネットワーク全体の繋がりの強さを示す指標であるEdge Betweennessを用いてネットワークによるクラスタリングを行った。

この手法をHUVEC細胞の時系列のTNF刺激実験に対して適用し、クラスタ抽出を行った。これらクラスタに対して遺伝子発現シグネチャー、GOからの用語関連遺伝子集合などと比較を行った。その結果、従来を知見と一致したセルサイクル、免疫機能、脂質代謝機能などの時系列の働きを知ることができた。