

遺伝子共発現の評価における発現レベル依存性

大林武

東北大学 大学院情報科学研究科

概要：遺伝子発現プロファイルの類似性である遺伝子共発現情報は、遺伝子機能を推定する手法として利用されており、我々の研究グループでは各生物ドメインを対象とした共発現データベース（ATTED-II, COXPRESdb, ALCOdb）の開発を行ってきた。遺伝子共発現情報の導出に当たっては、サンプル数が重要であることが知られており、公共のデータベースのデータの利用が有効である。しかし、サンプルの選択、発現データの補正、プロファイルの類似性の算出など、共発現関係を導出するためのパイプライン全体の最適化を行うには、ベンチマークデータに基づく評価が欠かせない。遺伝子共発現は定量的な関係性であるため、正解セットの作成は困難であるが、既知の遺伝子機能アノテーションとの無矛盾性に基づく評価が広く用いられている。ここで、既知の遺伝子機能アノテーションには、高発現遺伝子ほど多くのアノテーションが付いているという傾向がある。この傾向が高発現遺伝子の方が研究しやすいという研究バイアスなのか、高発現遺伝子は多くの機能を有するという真の傾向なのかを判断することは容易ではないが、もし研究バイアスだとすれば、このアノテーションに基づく共発現計算パイプラインの最適化は、高発現遺伝子の共発現性能の偏った最適化となってしまう。一方で、遺伝子共発現関係も発現レベル依存性が観察され、これが発現データに含まれる系統誤差、偶然誤差に基づくアーティファクトなのか、真の遺伝子関係を示しているものなのかの判断は難しい。本発表では、遺伝子共発現の性質や質を評価のために、遺伝子発現レベルがどのように関連するかを議論する。

キーワード：遺伝子発現、遺伝子共発現、遺伝子機能アノテーション