

# 大規模遺伝子発現量データを構成する 独立なサンプル数の極限值の見積もり

東北大学 大学院情報科学研究科

大林武

**概要：**RNAseqやマイクロアレイに基づく遺伝子発現量データは行列の形式で表現される。単一の実験を構成するサンプル数は数十程度のものが多いが、公共データベースに含まれる多数の実験を合わせることでヒト、マウス、ラット、シロイヌナズナなどのモデル生物種ではサンプル数が1万を超える遺伝子発現量行列を構築することができる。遺伝子間の発現プロファイル類似性である遺伝子共発現情報は、サンプル数に依存して情報の質が向上することが知られている。一方、サンプル数が多くなるにつれて類似環境のサンプルが多くなる。ここで、所与の遺伝子発現量データを構成する独立のサンプル数を見積もることができれば、それは発現データの質を評価する上で有用であるのみならず、独立サンプル数の極限は当該生命システムの複雑さを反映していると解釈でき、ゲノム進化の観点からも興味深い。本発表では、主成分分析による独立のサンプル数を見積もりとその種間比較について報告する。

**キーワード：**トランスクリプトーム、遺伝子共発現、遺伝子機能予測