

Omics 研究における検証的解析と探索的解析

松井 茂之

名古屋大学医学系研究科生物統計学分野

smatsui@med.nagoya-u.ac.jp

近年の著しいバイオ技術の進展により、生体由来の組織や細胞を分子レベルで捉えることが可能となり、DNA 配列（点変異，挿入・欠失，コピー数，染色体構造等），遺伝子発現，DNA メチル化，蛋白質発現，代謝物質などの様々な分子データの獲得が可能になっている。さらに，シーケンスやアレー技術の低コスト化により，全ゲノム，全遺伝子などを対象とした網羅的な解析の実施も比較的容易となり，分子データのビックデータ化が進んでいる。全ゲノム，全遺伝子などの分子全体を表すデータはオミクスデータ（omics data）と総称される。

オミクスデータと生体の表現型（罹患，病態，薬剤感受性など）の関連を調べることはオミクス研究の基本的な解析の一つである。具体的には，一つ一つのオミクス変数（遺伝子多型，プローブセットなど）と表現型との関連に対して検定が実施される形となる。オミクス変数の数は，数千～数百万とあるので，膨大な数の P 値が計算されることになる。関連解析において P 値をどう捉えるかは多重性調整のやり方に如実に表れる。これは，研究の性格や目的によって変わるが，基本的な考え方は，検証的解析と探索的解析で大きく分けることができる。本発表では，それぞれの解析における多重性調整について整理し，モデリングと推定の切り口から，検出力の評価や効率的な検定法などについても論じる。以上を踏まえ，最後にオミクス研究における P 値の役割についてまとめる。