

に対して適用する。

連続量として観測されるデータをカテゴリー化すると、一般に検出力の低下することが知られている。カテゴリー化を行う際には、検出したい効果を十分に検出できる検出力が確保されているかどうか確認しておくことが重要である。そこで、**Univariate test selection** を想定し、提案した方法の統計的な性能をシミュレーションによって評価する。

シミュレーションの結果、カテゴリー数 3 の場合、検討すべき遺伝子数を劇的に減らすことができるものの、検出力がほとんどなくなってしまうことが明らかとなった。また、遺伝子発現量変化が平均 0.3 以上ある遺伝子であれば、サンプル数 50 のカテゴリー数 5 で十分に検出可能であることが明らかとなった。

本研究が提案するカテゴリー化によるアプローチの特徴は、測定誤差を考慮しつつ個々の遺伝子の発現量変化をパターンに分類することができる点である。このため多くの遺伝子を少ないパターンに要約することができ、本研究で提案したアプローチのように、検討する変数の数を減らすことが可能になる。また他の特徴としては、ある閾値以上の変化をひとつのカテゴリーにするため、極端な外れ値に対して頑健であること、小さな遺伝子発現量変化をひとつのカテゴリーにするため、変化量の小さな遺伝子を除外できることが挙げられる。適用事例において明らかになったように、ある遺伝子の全ての対象者の変化量が小さな値であったとしても、結果変数との相関が強ければ統計的には有意となり選択されてしまう。このような遺伝子を選択したとしても、結果の再現性に疑問が残り、後続する臨床検査法開発の候補遺伝子としてはあまり役に立たないと思われる。

本研究で提案した **Resolution** は、カテゴリーの閾値を等間隔に取ったために、大変に簡便な式で表現することができた。本研究のような遺伝子発現量変化の場合、先行研究において頻用されるカテゴリー区分は存在しないため、カテゴリーを等間隔に設定しても構わないと思われる。**Resolution** は、連続量として測定される変数をカテゴリー化したときの指標であり、式が複雑になることを厭わなければ、等間隔でないカテゴリー区分にも対応可能であり、疫学研究における血圧測定など他分野へ応用することも可能であると考えられる。