

	きたむら のぶたか
氏名	北村 信隆
学位	博士 (医学)
学位記番号	新大院博(医)第234号
学位授与の日付	平成20年3月24日
学位授与の要件	学位規則第4条第1項該当
博士論文名	SNP データに基づく疾患感受性遺伝子同定のための多段階症例-対照研究デザインの特性
論文審査委員	主査 教授 赤澤 宏 平 副査 教授 山本 正 治 副査 教授 木南 凌

博士論文の要旨

【目的】

1 塩基多型 (single nucleotide polymorphism, SNP) やマイクロサテライトマーカ一等の DNA 多型データを用いた症例-対照関連解析は、疾患感受性遺伝子を見出す解析法として頻繁に用いられている。症例-対照関連解析の中でも疾患感受性遺伝子を効率的に検出するための研究デザインとして、多段階関連解析法があり、その代表的な手法として Replication-based analysis (RBA) と Joint analysis (JA) のふたつが知られている。多段階関連解析で用いられる研究デザインの多くは2段階法であるが、最近、極めて多数の候補遺伝子を対象として、3段階法を用いた疾患感受性遺伝子同定の研究が報告され始めた。しかし3段階法による疾患感受性遺伝子の統計学的検出力、陽性反応適中度 (PPV) およびタイピング数に関する研究は殆ど行われていない。

本研究の目的は、多段階関連解析法の特性を明らかにすることである。具体的には、実際の SNPs 解析で起こりうるさまざまな条件 (例えば、症例数、候補アリル数、実験全体の有意水準、各 k 段階 (k は 1 以上の整数) への症例数の配分割合 ($\pi_{s,k}$)、各 k 段階での候補アリルの選択割合 ($\pi_{m,k}$)、等) の下で、RBA と JA の統計学的検出力と PPV ならびにタイピング数の特性を調べた。

【方法】

プログラム言語として Mathematica ならびに R を用いて、RBA ならびに JA による検出力と PPV ならびにタイピング数を算出するプログラムを開発した。

【結果】

本研究から得られた結果は以下のとおりである。

1. 3段階法において、各ステージの症例割合を種々変化させたときの検出力とタイピング数を調べた。その結果、いずれの場合においてもJAがRBAよりも高い検出力を示し、例えば $\pi_{s,1}=0.4$ 、 $\pi_{s,2}=\pi_{s,3}=0.3$ の時、1段階法では、85.2%の検出力であるのに対して、3段階法のRBA、JAの検出力はそれぞれ49.9%、72.4%であり、JAの方が検出力の減少率は小さいことがわかった。その時のタイピング数は59.7%の削減が可能であった。

2. 最終ステージに残るアリの個数が一定数なることを仮定し、2段階法では、 $\pi_{m,1}=0.0001$ 、3段階法では、 $\pi_{m,1}=0.01$ 、 $\pi_{m,2}=0.01$ （すなわち $\pi_{m,1}\times\pi_{m,2}=0.0001$ となる）と設定して2段階法と3段階法とを比較した。その結果、例えば $\pi_{s,1}=0.2$ 、 $\pi_{s,2}=\pi_{s,3}=0.4$ のとき、2段階法のRBAとJAの検出力はそれぞれ14.8%、14.9%、3段階法のRBAとJAの検出力はそれぞれ46.6%、54.8%となり、同様に $0.1\leq\pi_{s,1}\leq 0.5$ において、RBAならびにJAによる3段階法の検出力ならびにPPVは2段階法よりも高い値を示す傾向が認められた。特に、 $\pi_{s,2}$ と $\pi_{s,3}$ が同じ程、またRBAよりもJAの方でその傾向が強くみられた。一方、 $0.6\leq\pi_{s,1}\leq 0.9$ の場合は、RBAならびにJAによる検出力とPPVは、3段階法よりも2段階法の方が常に高い値を示していた。

3. タイピング数を一定に設定した条件下において、各多段階デザインによる検出力ならびにPPVを比較したところ、3段階法と2段階法との検出力の差が小さくなっていったものの、最終段階で選択されるアリの数を一定とした条件の場合と同様、 $\pi_{s,1}$ が小さい範囲においては、RBAならびにJAによる3段階法の検出力ならびにPPVは2段階法よりも高い値を示す傾向が認められた。例えば $\pi_{s,1}=0.2$ 、 $\pi_{s,2}=0.3$ 、 $\pi_{s,3}=0.5$ のとき、2段階法のRBAとJAの検出力はそれぞれ45.9%、46.5%、3段階法のRBAとJAの検出力はそれぞれ46.0%、49.0%であった。

4. 症例数を増加させることにより各多段階デザインの検出力ならびにタイピング数も上昇したが、各デザインにおける上述のような変化のパターンの傾向は変わらなかった。なおJAにおいて、各症例数毎に多段階法の検出力とタイピング数を比較すると、いずれの症例数においても $\pi_{s,1}=0.3$ における3段階法の検出力は、 $\pi_{s,1}=0.5$ における2段階法の検出力とほぼ同程度の検出力を示していたが、前者のタイピング数は後者のタイピング数の約60%に抑えられていた。

【考察】

本研究により、2段階法だけでなく、 $0.1\leq\pi_{s,1}\leq 0.5$ においてJAを用いた場合には3段階法の有用性も示唆された。逆に $0.6\leq\pi_{s,1}\leq 0.9$ の場合は、RBAならびにJAによる検出力ならびにPPVは3段階法よりも2段階法の方が常に高い値を示していたことから、3段階法は用いるべきではないと考えられた。今後、SNPデータを用いた疾患感受性遺伝子の関連解析に際して、本プログラムにより、検出力推定のための事前シミュレーションを行うことが有用である。

(論文審査の要旨)

1 塩基多型 (以下 SNP と呼ぶ) データを用いた症例-対照関連解析は、疾患感受性遺伝子を見出す解析法として有用である。近年、対象症例を複数の層に分けて探索的解析と検証的解析を行なう多段階症例-対照関連解析が注目されている。

本研究の目的は、2つの多段階関連解析法 Replication-based analysis (RBA) と Joint analysis (JA) の特性を明らかにすることである。具体的には、実際の SNPs 解析で起こりうるさまざまな条件 (症例数、候補アリル数、実験全体の有意水準、各段階への症例数の割付け割合、各段階での候補アリルの選択割合) の下で、RBA と JA の統計学的検出力、陽性反応適中率 (PPV) ならびにタイピング数の特性を調べた。

3段階法におけるシミュレーションの結果、1) 各段階への症例数の割付け割合を変えても、JA は RBA に比べて検出力が常に高く、最大で 22.5% 高かった。また、タイピング数は 1段階法に比べて 59.7% の削減が可能であった。2) 第 1 ステージにおける症例数の割合が 0.5 以下のとき、3段階法の RBA、JA の検出力、PPV は、2段階法のそれらに比べて高い値を示した。

本論文は、研究者が制御可能なある条件下において、3段階の症例-対照関連解析が、2段階の場合に比べてより効率的に疾患関連遺伝子を同定できることを立証した点に、博士論文としての価値を認めた。