




審査結果の要旨

報告番号	甲 第 1256 号	氏名	佐藤俊太郎
審査担当者	主査	古川 恭治	
	副主査	吉浦 孝一郎	
	副主査	川口 淳	
主論文題目： Fast score test with global null estimation regardless of missing genotypes (欠測をともなう SNP-GWAS における高速なスコア検定法)			

審査結果の要旨 (意見)

全ゲノムシーケンス技術の著しい発展とともに、大規模遺伝情報の包括的な解析を効率的に行うことが求められている。本論文は、ゲノムワイド関連解析において、疾患発症の有無などの二値アウトカムと多数の SNP との関連を同時に調べる場合に役立つ高速スコア検定方法を提案し、各 SNP に欠測がある場合でも、推定性能をそれほど低下させることなく、計算コストを抑えることが可能となることを示している。従来法との比較における統計検出力の漸近的な一致性など、提案手法の数学的性質を注意深く導出しており、バイオ統計学の理論的研究としても価値を有する。よって、本論文の結果や内容は、博士号に値するものと評価できる。

論文要旨

Genome-wide association studies (GWAS) におけるケース・コントロール研究では、各 SNP が二値形質であるアウトカムに与える影響を検討する際に、性別、年齢等の交絡を除くためにロジスティック回帰モデルが利用される。各 SNP とアウトカムとの関連性は、Wald 検定、尤度比検定、またはスコア検定を SNP の数だけ繰り返し評価する。近年の GWAS において SNP の数は 50 万以上、全ゲノムシーケンシングでは数千万個のバリエーションを検定する必要があり、計算量の低減は重要である。

全 SNP かつ共変量に欠測がない場合、共変量を用いて計算する Null モデルが全 SNP で共通のため、Null 推定量を利用するスコア検定を用いれば計算負荷の大きいパラメータ推定が一度ですみ、Wald 検定や尤度比検定と比較して計算量を小さくできる。しかし、SNP データにはタイピングエラーによる欠測が通常含まれる。これにより、スコア検定であっても、SNP 毎に欠測がある個体を除外して Null モデル推定をする必要が生じ、計算量が増加してしまう。

本研究では、SNP に欠測があっても、全 SNP にわたって単一の Null モデルを利用するスコア検定を提案した。従来のスコア検定と提案法の検出力が漸的に一致することを証明し、シミュレーションにてタイプ I エラーおよび検出力を評価した。その後、GWAS 実データに提案法を適用し、従来の検定方法と比較して、提案法の計算速度が最大 18 倍速くなることを確認した。