

SNP 欠測を考慮した SNP-GWAS における 高速なスコア検定法

久留米大学バイオ統計センター，長崎大学病院 臨床研究センター
佐藤俊太郎
理化学研究所 革新知能統合研究センター
植木優夫

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，集団構造等の交絡を除くためにロジスティック回帰モデルがしばしば利用される．SNP とアウトカムとの関連性を見るために，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，各 SNP の影響を評価する．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，さらにこれを複数の形質値について検討することもあるため，計算量の低減は重要な課題である．

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はずである．しかしながら，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て，計算量が増加してしまうという問題がある．概要を図 1 に示す．

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 モデル推定量を利用するスコア検定を提案し，シミュレーションにてタイプ 1 エラーおよび検出力を評価する．

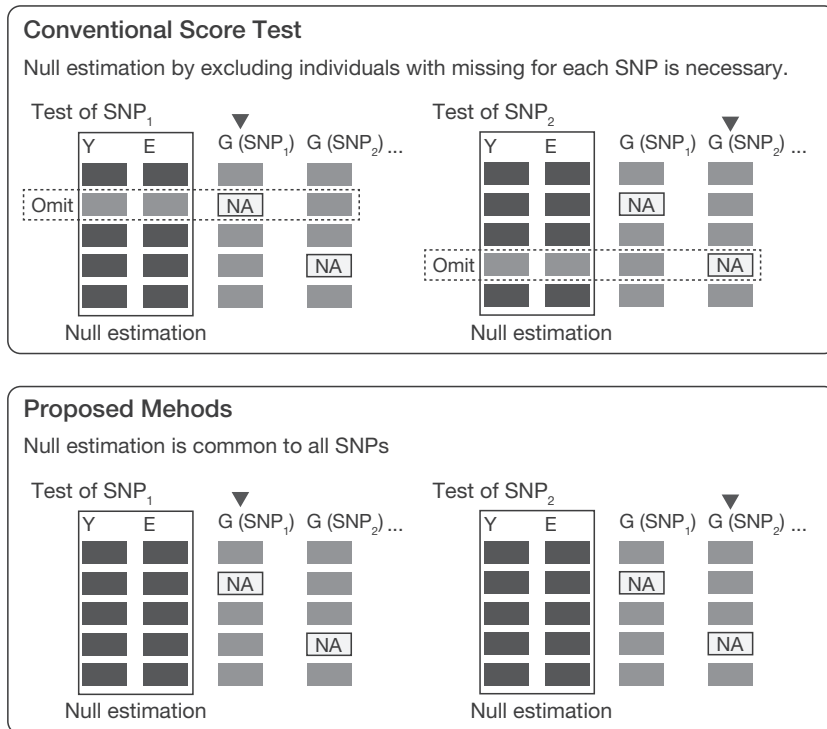


図 1: 研究の概要