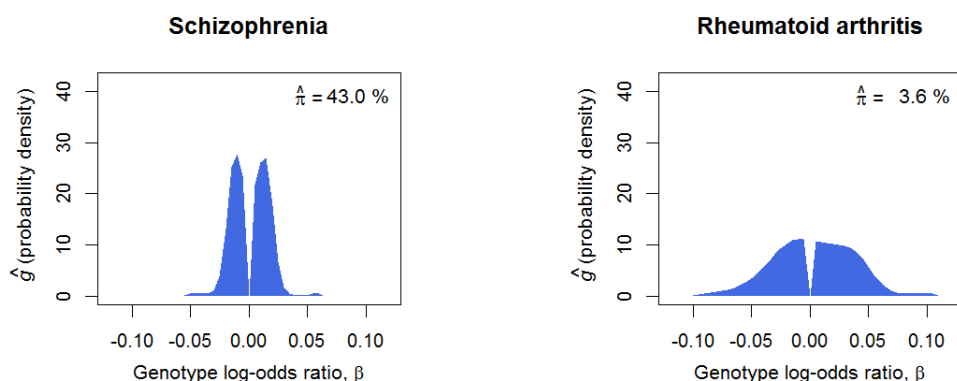


セミパラメトリック階層混合モデルを用いた多遺伝因子構造の推定

ゲノムワイド関連解析 (genome-wide association study; GWAS) は、ゲノム全体の塩基多型 (single nucleotide polymorphism; SNP) と形質との関連を網羅的に調べる方法であり、これまで多くの複雑疾患の疾患関連 SNP を特定してきました。一方で、これまで特定された疾患関連 SNP を集めても、家系から推定される複雑疾患の遺伝性の割合 (遺伝率) はごく一部しか説明できないことも分かってきました。この問題は「失われた遺伝率」 (missing heritability) と呼ばれ、複雑疾患の遺伝学的な発生機序の解明やそれに基づいた医療の大きな障壁となっています。

「失われた遺伝率」を説明するひとつの仮説は、疾患関連 SNP の効果サイズが小さいために現在のサンプルサイズでは検出できていない、というものです。しかし、これらのゲノム全体に存在する小さい効果も含めた疾患関連 SNP の割合や効果サイズ分布を柔軟にかつ容易に定量できる手法は存在しませんでした。そこで我々は、これまで遺伝子発現解析で開発されてきたセミパラメトリック階層混合モデルに注目し、これを GWAS 用に発展させました [1]。本手法は、しばしばウェブ上で公開されている SNP の要約統計量のみを使用し、少ない計算資源と時間で実行可能です。

本手法を、統合失調症 GWAS データに適用したところ、ゲノム全体の独立な SNP の中で疾患関連 SNP の割合 (π) は 40% にも及ぶ一方でそれらのリスク遺伝子型の効果サイズ分布 (g) の大部分はオッズ比にしてわずか 1.03 に収まると推定されました。これとは対称的に、関節性リウマチデータでは、4%程度がリスク変異であるが、それらの効果サイズ分布は、オッズ比にして 1.05~1.1 と比較的に効果の大きい変異が多いことが分かりました。また、末梢血において遺伝子発現を調整する SNP は、関節性リウマチにおいては、大きな疾患関連性を有するのに対して、統合失調症では他の SNP と大きな違いがないことも分かりました。このように、セミパラメトリック階層混合モデルは、複雑疾患の多遺伝因子構造の理解を促し「失われた遺伝率」の解明の一助となると考えられます。



1. Nishino J, Kochi Y, Shigemizu D, Kato M, Ikari K, Ochi H, Noma H, Matsui K, Morizono T, Boroevich K.A, Tsunoda T, Matsui S. (2016). Empirical Bayes estimation of semi-parametric hierarchical mixture models for unbiased characterization of polygenic disease architectures. bioRxiv doi.org/10.1101/080945