多次元階層混合モデルとoptimal discovery procedureによる遺伝子-治療交互作用の検出

大谷 隆浩 リスク解析戦略研究センター 特任助教

はじめに

特定の疾患における遺伝子と治療の交互作用を検出することは，個別化医療の実現に向けた重要な課題であり，数多くの研究がこれまでに行われている．しかしながら，ゲノム全体を対象とした網羅的な解析を行う場合，交互作用項を取り入れた回帰モデルに基づく検定の検出力は低く，その多くが見落とされている可能性がある．これに対して，近年，Matsui et al.[1]によりセミパラメトリック階層混合モデル[2,3]とoptimal discovery procedure (ODP)[4,5] を用いた遺伝子スクリーニング手法が提案され，遺伝子発現マイクロアレイ解析が付随したランダム化臨床試験の実データ解析とシミュレーション実験により，既存の検定手法に比べて高い検出力を達成できることが示されている．

本研究では，この解析手法を一塩基多型 (single nucleotide polymorphism; SNP) を対象としたゲノムワイド関連解析 (genome-wide association study; GWAS) に応用し，SNPと治療の交互作用の検出を試みた．具体的な事例として，脳梗塞の再発を主要エンドポイントとした，マルチビタミン剤の高用量摂取に関する多施設共同二重盲検ランダム化比較試験「VISP (Vitamin Intervention Stroke Prevention) trial[6]」のGWASデータセットに注目し，これに対して上記の方法による再解析を行った．

解析方法

Matsui et al.[1]によって提案されている，遺伝子-治療交互作用の検出手順を**図1**に示す．この手法では，まず対象者を対照群と治療群に分け，それぞれの群で独立に回帰モデルによる関連解析を行う．これにより，対照群におけるSNP主効果の効果量の推定値$b_j^{(c)}$と，治療群における推定値$b_j^{(t)}$が求められる $(j=1,\dots,M)$．そして，これらの効果量$\boldsymbol{b}_j=(b_j^{(c)},b_j^{(t)})$の潜在的な分布$f$として，以下の階層混合モデル[2,3]を仮定する．

$$f(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j) = \underbrace{\pi f_0(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j)}_{\text{Null}} + \underbrace{(1-\pi) f_1(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j)}_{\text{Non-null}}$$

$$\boldsymbol{b}_j \sim N(\boldsymbol{0},\boldsymbol{\Sigma}_j) \qquad \boldsymbol{b}_j|\boldsymbol{\beta}_j \sim N(\boldsymbol{\beta}_j,\boldsymbol{\Sigma}_j),\ \boldsymbol{\beta}_j \sim g_1(\cdot)$$

ここで，$\boldsymbol{\Sigma}_j=\mathrm{diag}\left(V_j^{(c)},V_j^{(t)}\right)$は$\boldsymbol{b}_j$の推定分散を対角とする分散共分散行列であり，$\pi$はSNPの効果量が$\boldsymbol{0}$ (疾患に関連しない) となる事前確率を表す．f_0は疾患に関連しないSNPに対する$\boldsymbol{b}_j$の分布を表し，f_1は効果量が非$\boldsymbol{0}$の (疾患に関連する) SNPに対する分布を表す． f_1には階層モデルを仮定し，効果量分布g_1は任意の分布を設定できる．ここではg_1についてノンパラメトリックな分布を仮定し，πとともに，EMアルゴリズムによってデータに基づいて推定する(経験ベイズ推定)．この推定結果に基づいて，各SNPについて疾患関連の事後確率を求めることができる．

さらにこの推定分布と事後確率を用いて，ODP[4,5]により疾患関連SNPを検出する．この手法において，各SNPの順位付けを決める統計量は$R_{\mathrm{ODP}}(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j)=\hat{f}_1(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j)/f_0(\boldsymbol{b}_j,\boldsymbol{\Sigma}_j)$であり，この統計量が大きいSNPを優先的に抽出する．抽出したSNP群に対するfalse discovery rate (FDR)は，各SNPの疾患関連の事後確率から推定できる．

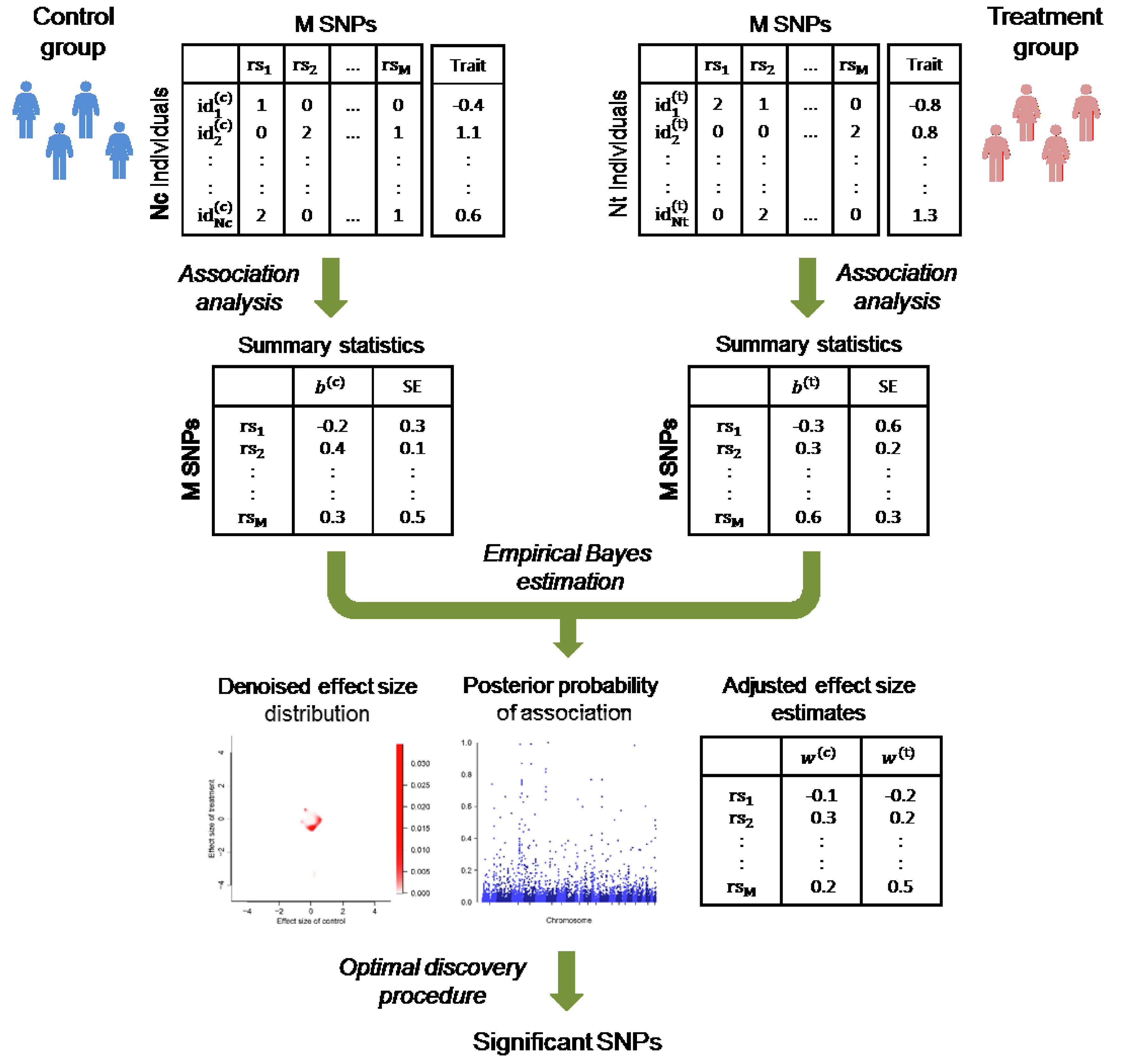

図1 階層混合モデルとODPによる遺伝型-治療法交互作用の検出手順．

解析事例

本研究では，VISP trial[6]のデータセットを例に解析を行った．この試験は，葉酸，ビタミンB6，B12を含むマルチビタミン剤を高用量日常摂取することで，脳梗塞の再発や非致死性心筋梗塞，死亡率を低下させることができるか検証することを目的に実施された．治療群では，25mgのピリドキシン(B6)，0.4mgのコバラミン(B12)，2.5mgの葉酸を摂取し，対照群ではそれぞれ200mcg，6mcg，20mcgを摂取した．対象者数は2164人であり，Illumina HumanOmni1-Quad_v1-0_Bにより約100万のSNPについて各個人の遺伝型が決定された．

ここでは従来の検定手法との比較のために，Wakefield et al.[7]によって行われた解析に基いて再解析した．この解析では簡単のために，ベースライン時と初回観察時での血中ホモシステイン濃度の差をアウトカムとしている(一般的に，血中ホモシステイン濃度が高いと循環器疾患を引き起こしやすいと言われている)．

Wakefield et al.は，交互作用項を含む一般的な線形回帰モデルに基づいた解析により，ビタミン剤摂取との交互作用を持つ3つのSNPを同定している[7]．本研究では，まず，階層混合モデルの経験ベイズ推定により，疾患関連SNPの割合と，その効果量の分布を推定した(**図2**)．推定された分布からわかるように，多くの疾患関連SNPの効果量は両群において小さいと考えられる．しかしながら，別のピークが$(b_j^{(c)},b_j^{(t)})=(0.0,-3.3)$の位置にあり，これは特定のSNPとビタミン剤高用量摂取に，効果量の大きい交互作用があることを示している．

次に，この効果量分布の推定結果に基づいて，各SNPについて疾患関連の事後確率を求めた(**図3**)．さらに，これをもとにODPを用いて，FDR=5%の基準のもとで，疾患に関連すると考えられるSNPを検出した．検出されたSNPの中にはWekefield et al.が報告しているものも含まれるが，連鎖不平衡を考慮した上で，新たに3つのSNPを検出できた．

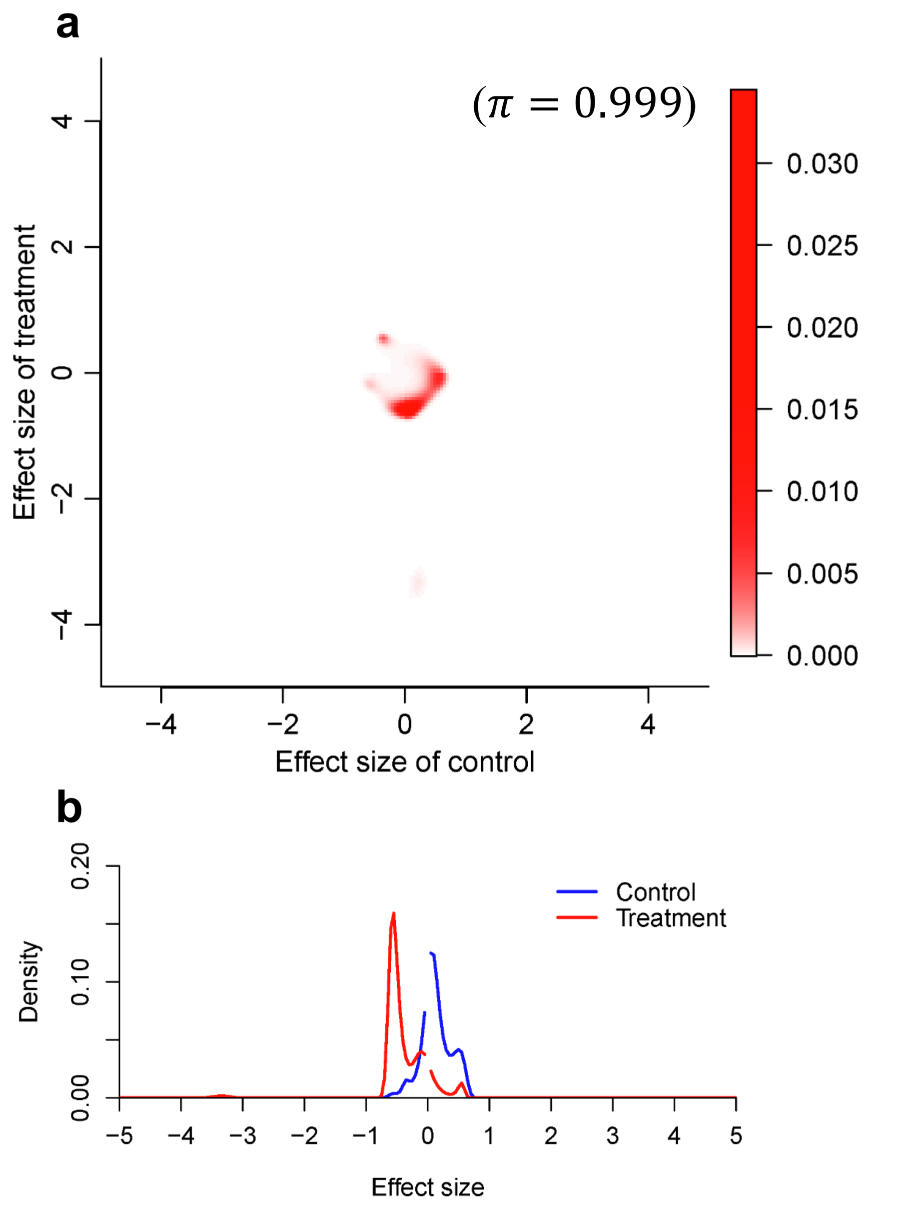

図2 疾患関連SNPの効果量の分布．
(**a**) 対照群・治療群での二次元分布．
(**b**) 各群の周辺分布．

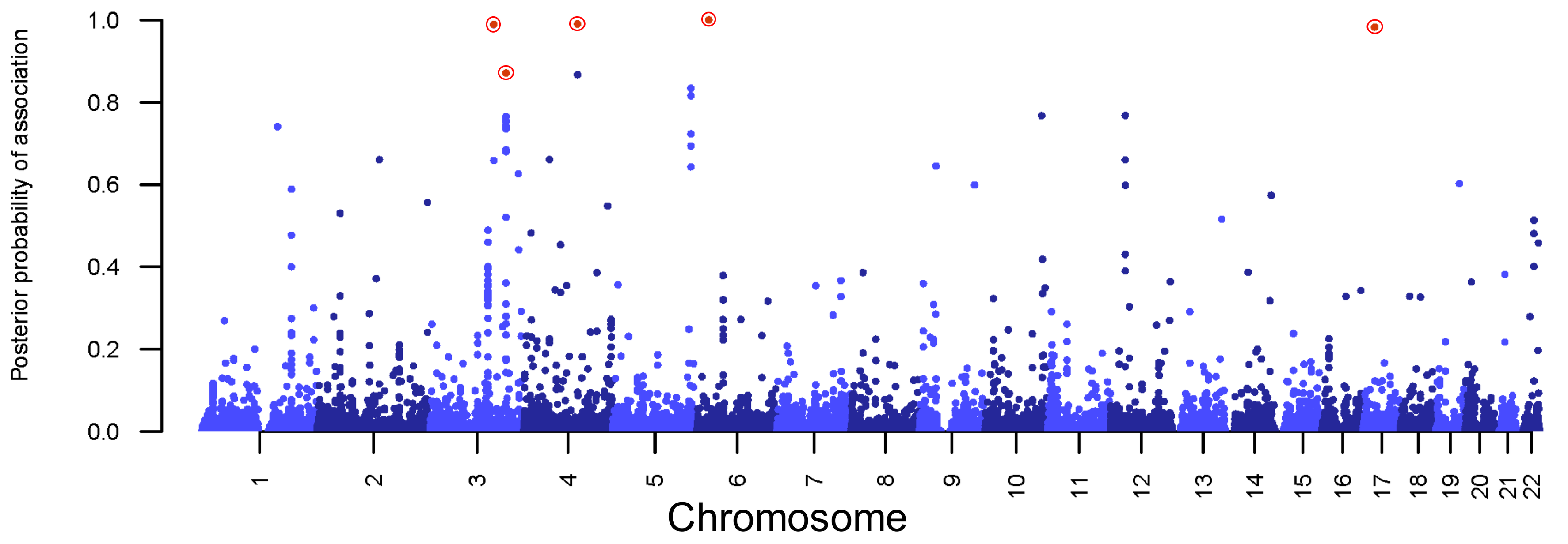

図3 各SNPに対する，疾患関連の事後確率のプロット．FDR=5%の基準のもとで有意となったSNPのうち，連鎖不平衡にない独立なものを二重丸で示した．

また，交互作用項を取り入れた線形回帰モデルとマイクロアレイ実験などで一般的に用いられる検定手法であるqvalue[8]を用いた解析と，階層混合モデルとODPの組み合わせによる解析での，有意となるSNPの数を比較した(**表1**)．偽陽性の可能性があるため単純な評価はできないものの，ODPでは同一のFDR基準のもとで，より多くのSNPを検出している．

表1 各FDR基準における，有意となるSNPの数の比較．なお，連鎖不平衡にあるSNPを含むため，独立なSNPの数はより少ない．

FDR	qvalue	ODP
1%	0	5
5%	1	8
10%	1	12
20%	2	26

まとめ

マルチビタミン剤の高用量摂取に関するランダム化比較試験のデータセットに対して，階層混合モデルとODPを用いた再解析により，SNP-治療交互作用の検出を試みた．その結果，循環器疾患に影響する血中ホモシステイン濃度を大きく低下させる効果を持つ交互作用を，新たに3つ検出した．より詳細な検証は必要なものの，これらは個別化医療における有用なバイオマーカーとして利用できる可能性がある．

参考文献

[1] Matsui et al., 2017; Biometrics (DOI: 10.1111/biom.12716).
[2] Matsui and Noma, 2011; Biometrics 67: 1225-1235.
[3] Nishino et al., 2016; bioRxiv (DOI: 10.1101/080945).
[4] Storey, 2007; J Royal Statist Soc B 69: 347-368.
[5] Noma and Matsui, 2012; Statist Med 31: 165-176.
[6] Toole, 2002; J Neurol Sci 15: 203-204 (dbGaP Study Accession: phs000343.v3.p1).
[7] Wakefield et al., 2014; Pharmacogenomics J 14(4): 309-315.
[8] Storey, 2002; J R Statist Soc B 64: 479-498.