

論文の内容の要旨

論文題目 バイズ流アプローチを用いたゲノムワイド関連研究におけるマーカー選択

指導教員 大橋靖雄 教授

東京大学大学院医学系研究科

平成 17 年 4 月進学

保健学博士課程

健康科学・看護学専攻

口羽 文

緒言

ヒトゲノム配列の解読が終了したことにより、集団中にありふれていて、より複雑な疾患に関する遺伝要因の研究が盛んに行われるようになった。多くの「ありふれた」疾患は、複数の遺伝子、環境要因、また、それらの交互作用により引き起こされることが知られている。そのため、遺伝要因の解明にはゲノム全体を対象とするアプローチが強力な手段になり得ることが指摘されていた。近年では、HapMap プロジェクトにより、ヒトゲノム上の遺伝的変異パターンは急速に解明されつつある。また、遺伝子型測定技術の進歩も著しく、近年では、数十万個の SNP を同時に扱うことができるようになった。これらのことから、「ありふれた」疾患に関する疾患感受性遺伝子の特定に大きな期待が持たれている。しかし、疾患と真に関連があり生物学的にも重要な遺伝子でもその効果は比較的小さいことが予想されており、疾患とは無関係な大部分の SNP との分離は困難なものとなっている。そのため、このような超大量データを扱う研究デザインや解析方法のさらなる発展が必要とされ

ている。

ゲノムワイド関連研究では、コスト効率の観点より、すでに多段階デザインを用いることが一般的であるが、どのような SNP のサブグループを追跡することが最適であるか、という問題に対して、多くは統計学的有意性を用いて検討されてきた。この方法は、研究対象に対して特に事前情報がなく、すべての SNP を等価として考えることが妥当である場合には有用な方法である。一方で、すでにヒトゲノムに関する多くの研究がおこなわれており、これらより得られる事前情報を考慮して SNP 選択を行う方法が提案されつつある。事前情報を積極的に用いることのできる方法の 1 つに各 SNP に対してベイズ流の推論を行うことが考えられる。ベイズ流の推論とは、既存の知識を反映する事前分布とデータより得られる尤度の両方を用いて得られる事後分布に基づいて推論を行う方法である。広範囲の事前情報を考慮できることや推定量の性質のよさより、遺伝子研究の分野においても注目されつつある。そこで、本論文では、ゲノムワイド関連研究 1 次スクリーニングにおける SNP 選択方法として、さらに検討すべき SNP の優先順位を付けるという観点より、連鎖不平衡パターン、事前情報を柔軟に考慮できる選択方法を提案する。また、本論文の動機付けとなったケース・コントロール研究デザインによる肺腺がんゲノムワイド関連研究データに基づくシミュレーションを行うことによって、その性能を評価する。さらに、標準的な方法と提案する方法の両方を実データに適応した結果の比較検討を行う。

方法

$j(j=1, \dots, m)$ 番目の SNP に対する統計量 T_j は、平均 θ_j 、分散 1 の正規分布に従うとする。

標準的な方法では、この T_j の大きさに基づいて SNP が選択される。これは p 値による選択と等価である。

ゲノムワイド関連研究では、疾患と関連のある SNP が少数存在し、そのほかの大部分の SNP は

疾患とは関連がないという事前情報が存在する. この事前情報を個々の SNP に考慮するため, θ_j に対して, 以下の事前分布を考える.

ここで, は, 0 の point mass を持つ関数である. 第 1 項が関連ありの場合, 第 2 項が関連なしの場合に θ_j が従うと考えられる分布を表している. P は, SNP $_j$ が真に疾患と関連がある確率であり, このときは平均 μ , 分散 τ_j^2 の正規分布に従う. また, SNP $_j$ は確率 $1-P$ で疾患と関連がなく, その場合には $\theta_j=0$ の質点を持つとする. これらより, 事後期待値 $E[\theta_j|T_j]$ は,

と求めることができる. ただし,

である.

事後期待値は, 疾患と関連があった場合の事後期待値にさらに“データを与えた下で効果がある確率”と解釈できる w_j が考慮された指標であると考えられる. この事後期待値に基づき SNP 選択を行う. P , μ は外部情報より事前に設定できるものとする. また, τ_j^2 は経験ベイズ法により観測データを用いて最尤推定することを考える. このことにより, 各 SNP に対する推論に観測データの連鎖不平衡パターンを反映させることができる.

シミュレーションによる評価と結果

標準的な方法による SNP 選択と事後期待値による方法の性能を比較検討するため, シミュレー

ションを行った。本論文の動機付けとなった肺腺がんゲノムワイド関連研究データの連鎖不平衡パターンを反映させるため、実データのコントロール群を用いて擬似データを発生させた。1,000SNP中、真に疾患と関連のあるSNP(以下、感受性SNP)が1つ存在するとし、この感受性SNPの位置を変えることによって、その周辺の連鎖不平衡パターンをいくつか設定した。感受性SNPの位置を8通り、オッズ比=1.2, 1.5の2通りの合計16通りの擬似ケース・コントロール集団における検討を行った。事前情報を反映した事前分布として、 $P=0.001$, $\mu=1, 2, 3$ を設定し、感度・特異度、感受性SNPの平均的な順位、選択されるSNPの感受性SNPからの距離の3つの観点より評価を行った。

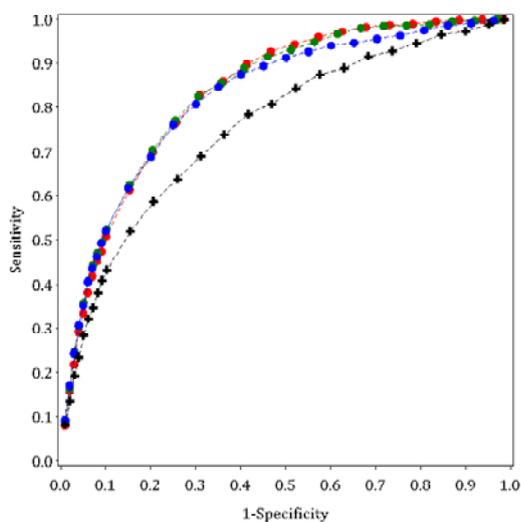
図1にROC曲線を示す。ROC曲線は、選択基準とした上位S%を $S=0, 1, 2, \dots, 10, 15, \dots, 100(\%)$ と変化させ、選択基準ごとに感度、特異度を求めることによって作成した。 μ の設定に依存するものの、標準的な方法に比べて、感度、特異度の観点より性能が改善されることが示された。また、同様の状況設定において、標準的な方法を用いた場合の感受性SNPの順位は233.6であった。一方、 $\mu=1, 2, 3$ における事後期待値による順位は、それぞれ160.6, 161.8, 175.6となり、感受性SNPの優先順位が上昇することが示された。さらに、図2に、上位50番目までのSNPと感受性SNPとの距離の関係を表す。特に連鎖不平衡の強い領域内に感受性SNPが存在している場合には、事後期待値を用いる方がより感受性SNPに近いSNPが上位に選択される傾向にあることが示唆された。

また、肺腺がんのゲノムワイド関連に適応した結果、標準的な方法を用いた場合と比べ、優先順位が異なるSNPが存在することが確認された。 $\mu=2$ において、どちらかの方法により5,000番以内に含まれたSNPの順位をプロットした結果を図3に示す。

考察と結論

ゲノムワイド関連研究において、さらに検討すべきSNPの最適な選択方法を探索することを目

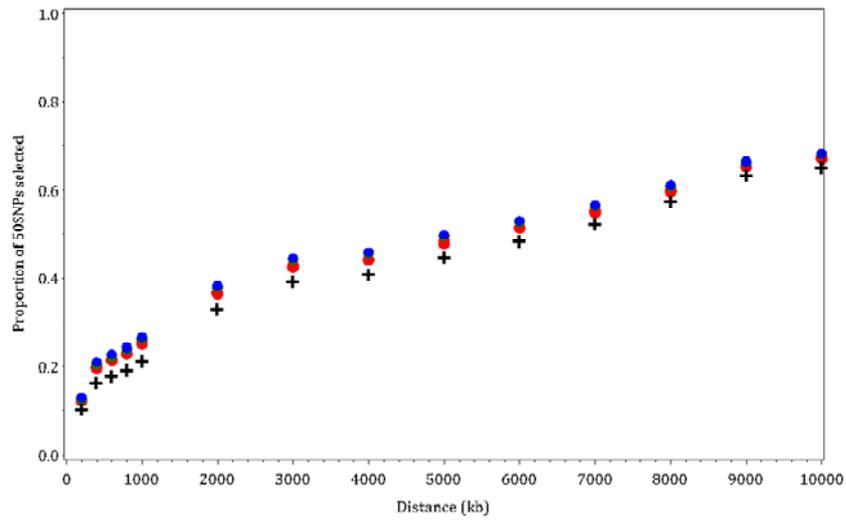
的とし、本論文では、既存の知識と観測データの連鎖不平衡パターンを考慮できる事後期待値による選択方法を提案した。ベイズ流のアプローチを用いることで、各 SNP に対する事前情報をより柔軟に利用することが可能となる。また、ハイパーパラメータの設定により、戦略的な SNP 選択が可能になると考えられる。シミュレーションによる評価の結果、標準的に用いられている方法と比べ、同等かそれ以上の性能をもつことが示された。また、ゲノムワイド関連研究では、膨大な数の SNP に対する検討が必要であるが、提案する方法は計算負荷が少ないため、容易に適応可能であることも利点であると考えられる。しかし、性能の改善の程度は、事前分布の設定や感受性 SNP が存在する領域の連鎖不平衡パターンに依存する。近年では、各 SNP に対するより詳細な特徴が比較的容易に得られるようになりつつある。よって、1 次スクリーニングのように大規模で、より探索的な SNP 選択において、これらの情報の最適な利用法にはさらなる検討が必要であると考えている。



Statistics ●●● PosteriorE(Myu=1) ●●● PosteriorE(Myu=2) ●●● PosteriorE(Myu=3) +++ Standard

図 1. ROC 曲線.

縦軸は感度、横軸は 1-特異度を表す。Standard は標準的な方法、PosteriorE(Myu=a)は $\mu = a$ としたときの事後期待値を示す。



Statistics ●●● PosteriorE(Myu=1) ●●● PosteriorE(Myu=2) ●●● PosteriorE(Myu=3) +++ Standard

図 2. 選択された SNP と感受性 SNP との距離.

縦軸は感受性 SNP とある一定の距離 (kb) 以内に含まれる SNP の割合, 横軸は感受性 SNP との距離を表す. Standard は標準的な方法, PosteriorE(Myu=a)は $\mu = a$ としたときの事後期待値を示す.

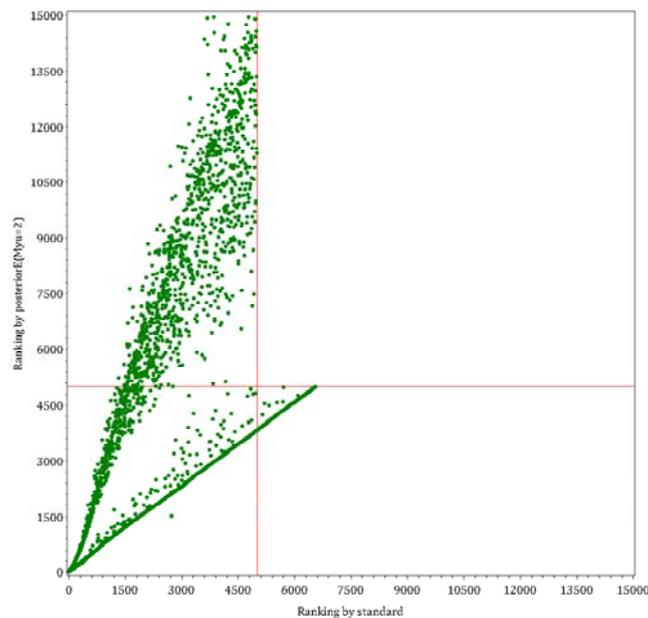


図 3. 標準的な方法あるいは事後期待値による方法のどちらかの方法にて上位 5000 番目に入る SNP の順位の比較.

縦軸が事後期待値 ($\mu = 2$) による順位, 横軸が標準的な方法による順位を表す. 優先順位が高い順に 1 番から順位をつけた. 赤線は 5000 番目を示す.