

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：37104

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25870074

研究課題名(和文) 強力なゲノムワイド遺伝子×遺伝子および遺伝子×環境相互作用解析法の開発

研究課題名(英文) Development of powerful genome-wide gene-gene and gene-environment interaction analysis

研究代表者

植木 優夫 (Ueki, Masao)

久留米大学・バイオ統計センター・准教授

研究者番号：10515860

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：ゲノムワイド関連解析(GWAS)における一塩基多型(SNP)データに対し、遺伝子×遺伝子相互作用、あるいは遺伝子×環境相互作用解析のための統計手法は十分に整備されていない。本課題においては、相互作用の検定において、自由度を分割表の観察頻度に応じて適応的に変化させる強力な検定手法を開発した。さらに、遺伝的予測問題において、次世代シーケンシングデータのような大規模データであっても迅速に実行可能で、なおかつ高精度なスパースモデリング手法、smooth-threshold multivariate genetic prediction (STMGP)法を開発した。

研究成果の概要(英文)：World-wide genome-wide association studies (GWASs) using single nucleotide polymorphisms (SNPs) have revealed that SNPs explain small portion of variability of human phenotypes, so-called the missing heritability problem. Current standard method which tests each SNP one at a time may miss contributions by gene-gene or gene-environment interactions. However, statistical methods for those hypotheses are still underdeveloped. In this project, a new statistical testing method is developed. It allows to vary the degrees of freedom for test statistics according to the observed cell frequencies. Simulation studies confirmed that a practically adequate statistical power is achieved. In addition, for genetic prediction problem, new sparse modeling method, smooth-threshold multivariate genetic prediction (STMGP), is developed. It is very rapid even in application to large-scale data such as whole-genome sequencing data, and is shown to have higher prediction accuracy.

研究分野：遺伝疫学、統計遺伝学

キーワード：遺伝的相互作用解析 ゲノムワイド関連解析

1. 研究開始当初の背景

ヒト疾患において、失われた遺伝率問題を説明するためのひとつの候補として、遺伝子×遺伝子あるいは遺伝子×環境相互作用が挙げられていた。これらの及ぼす影響の調査と解析は、単一 SNP(一塩基多型)遺伝子関連解析に代わる次世代の研究対象となっていた。網羅的なゲノムワイド探索を実施するための方法論の構築が必要とされていたが、莫大な数の SNP を対象とした網羅的探索は計算コストの問題が深刻であり、さらには、検定の多重性、高次元の局外パラメータ、データのスパース性、等の統計的な問題から生じるパワー損失へも懸念があった。

2. 研究の目的

本研究の目的は、相互作用解析における統計的課題に取り組むことで、大幅なパワー向上を達成する、新たなゲノムワイド相互作用解析法を開発することである。その上で、失われた遺伝率問題を読み解くための知見を見出すことを狙いとしている。

3. 研究の方法

これまで研究代表者が開発した遺伝子×遺伝子相互作用検定法と遺伝子×遺伝子相互作用変数選択法を発展させつつ、統計手法の改良を理論面から考察し、新たな手法開発を試みた。生物学的、遺伝学的に妥当な数理モデルに基づいた数値実験によって、現存する統計手法との比較検証を行った。公開されている実際のゲノムワイド SNP データを用いた適用を行った。

4. 研究成果

SNP×SNP データに対する様々な仮想モデルから発生させた人工データを用いて、BOOS

T法(Wan et al. 2010, Am J Hum Genet)の挙動をより詳細に調べた。その結果として、検出力を向上させる方法を新たに開発できた。具体的には、現実的なサンプル数のもとで頻繁に遺伝子型の分割表に空セルが生じるが、その場合に、従来の固定自由度の検定では検定が保守的となることを見出し、その代わりに自由度を適応的に変動させることで検出力を高めるといった単純な方法である。提案手法は既存のBOOSTプログラム等を利用することで容易に実行可能である。人工データを用いた数値実験を通じて、提案手法が実際に検出力を高めることが確認できた。得られた成果を論文として出版し(Ueki 2014, Stat Med)、また、シンポジウムおよび講演にて研究発表を行った。一方、最新の遺伝統計手法をレビューした内容をまとめて、書籍を一部執筆した。その後、遺伝的予測問題の検討に着手した。予測モデルの構築は個別化医療の実現に必須であるが、失われた遺伝率問題により、ゲノムワイド SNP を用いても予測精度が充分に得られないことが多々報告されていた。

SNP データは超高次元であり、統計学分野においては、 $p \gg n$ 問題と呼ばれる困難が存在している。過去に研究代表者が開発した統計手法を拡張することで予測精度の向上が図れるものと予想し、SNP-GWAS を用いた高速な予測手法である STMGP(smooth-threshold multivariate genetic prediction)法を新たに開発することができた。数値実験ならびに米国 Alzheimer's Disease Neuroimaging Initiative(ADNI)が提供する実際の次世代シーケンシングデータへの適用を通じて、既存手法に比べて高い予測精度をもつことが確かめられた。得られた成果は国際誌に掲載され(Ueki and Tamiya 2016, Genet Epidemiol)、シンポジウムおよび講演にて研究成果の発表を行った。さらに、研究代表者が一部データ解析を担当した共同研究が論文として出版された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 10 件)

1. Ueki M, Tamiya G. Smooth-threshold multivariate genetic prediction with unbiased model selection. **Genet Epidemiol** 40:233-243, 2016. doi:10.1002/gepi.21958.
2. Kawasaki Y, Ueki M. Sparse predictive modeling for bank telemarketing success using smooth-threshold estimating equations. **J Jpn Comp Statist** 28:53-66, 2015.
3. Okamura K, Ohe R, Abe Y, Ueki M, Hozumi Y, Tamiya G, Matsunaga K, Yamakawa M, Suzuki T. Immunohistopathological analysis of frizzled-4-positive immature melanocytes from hair follicles of patients with Rhododendrol-induced leukoderma. **J Dermat Sci** 80:156-8, 2015.
4. Ueki M. On the choice of degrees of freedom for testing gene-gene interactions. **Stat Med**. 2014 Dec

- 10:33(28):4934-48. doi:
10.1002/sim.6264. Epub 2014 Jul 10.
5. Shimanuki M, Abe Y, Tamiya G, Ueki M, Hozumi Y, Suzuki T. Positive selection with diversity in oculocutaneous albinisms type 2 gene (OCA2) among Japanese. **Pigment Cell Melanoma Res.** 2015 Dec 3 28:233-235. doi: 10.1111/pcmr.12337.
 6. Tamiya G, Makino S, Hayashi M, Abe A, Numakura C, Ueki M, Tanaka A, Ito C, Toshimori K, Ogawa N, Terashima T, Maegawa H, Yanagisawa D, Tooyama I, Tada M, Onodera O, Hayasaka K. A mutation of COX6A1 causes a recessive axonal or mixed form of Charcot-Marie-Tooth disease. **Am J Hum Genet.** 2014 Sep 4;95(3):294-300. doi: 10.1016/j.ajhg.2014.07.013.
 7. Yamagata University Genomic Cohort Consortium (YUGCC). Pleiotropic effect of common variants at ABO Glycosyltransferase locus in 9q32 on plasma levels of pancreatic lipase and angiotensin converting enzyme. **PLoS One.** 2014 Feb 28;9(2):e55903. doi: 10.1371/journal.pone.0055903. eCollection 2014.
 8. Iwano M, Igarashi M, Tarutani Y, Kaothien-Nakayama P, Nakayama H, Moriyama H, Yakabe R, Entani T, Shimosato-Asano H, Ueki M, Tamiya G, Takayama S. A pollen coat-inducible autoinhibited Ca²⁺-ATPase expressed in stigmatic papilla cells is required for compatible pollination in the Brassicaceae. **Plant Cell.** 2014 Feb;26(2):636-49.
 9. Ueki M, Kawasaki Y. Multiple choice from competing regression models under multicollinearity based on standardized update. **Comput Stat Data Anal.** 2013;63:31-41.
 10. Shibata K, Hozawa A, Tamiya G, Ueki M, Nakamura T, Narimatsu H, Kubota I, Ueno Y, Kato T, Yamashita H, Fukao A, Kayama T; Yamagata University Genomic Cohort Consortium. The confounding effect of cryptic relatedness for environmental risks of systolic blood pressure on cohort studies. **Mol Genet Genomic Med.** 2013 May;1(1):45-53.
- 〔学会発表〕(計 15 件)
1. 植木優夫, 田宮元. 円滑閾値型推定方程式による遺伝的予測. 科研費シンポジウム: 生命科学データ解析の方法論と健康科学への応用. 東京大学. 東京. 2015 年 10 月 16 日.
 2. 植木優夫, 田宮元. 円滑閾値型推定方程式による遺伝的予測. 2015 年度 統計関連学会連合大会 岡山大学. 岡山. 2015 年 9 月 9 日.
 3. 植木優夫. ゲノムワイドな遺伝子 x 遺伝子間相互作用の探索 . 科学研究費・基盤研究(A)「非対称・非線形統計理論と経済・生体科学への応用」シンポジウム 「バイオ統計学の挑戦と貢献」. 福岡; 2015 年 2 月.
 4. 植木優夫. 遺伝子 x 遺伝子間相互作用検定の自由度について . 2014 年度統計関連学会連合大会 . 東京; 2014 年 9 月.
 5. 植木優夫, 川崎能典. 多重共線条件

- 下における複数の回帰モデル選択 . 2013 年度統計関連学会連合大会 . 大阪; 2013 年 9 月.
6. 植木優夫, 嶋村海人, 川野秀一, 小西貞則, 田宮元. 超高次元スパース回帰法によるゲノムデータ解析. 岡山大学 IPSR × 九州大学 IMI × 理研 CSRS シンポジウム 生命データ科学による新たな社会的価値の創造 ~ 医療、農業、環境分野における役割と作物設計への応用 ~. 理化学研究所 横浜キャンパス. 神奈川. 2016 年 2 月 23 日.
 7. 植木優夫. ヒトゲノムデータから見た統計科学. 遺伝学と統計学における数理とモデリング. 政策研究大学院大学. 東京 2016 年 1 月 25 日.
 8. Ueki M, Tamiya G. Multivariate smooth-threshold genetic prediction with unbiased model selection. East Asia Regional Biometric Conference 2015, Fukuoka. 2015 年 12 月 22 日.
 9. Ueki M, Tamiya G. Smooth-threshold multivariate genetic prediction with unbiased model selection. 広島統計談話会. 放射線影響研究所. 広島. 2015 年 12 月 4 日.
 10. Ueki M, Kawasaki Y. Choosing multiple good regression models under multicollinearity, CSA-KSS-JSS joint International Session. Taipei; December 2014.
 11. 植木優夫:ゲノムワイド疾患関連解析における統計的諸問題. 第 13 回久留米大学バイオ統計学フォーラム. 福岡 2014 年 9 月
 12. Ueki M, Tamiya G. Ultrahigh-dimensional variable selection for genome-wide SNP-SNP interaction analysis. The 3rd Institute of Mathematical Statistics Asia Pacific Rim Meeting. Taipei; 2014 June-July.
 13. 植木優夫 : ゲノムワイド SNP-SNP 相互作用解析. 2013 年度統計関連学会連合大会. 大阪; 2013 年 9 月.
 14. Ueki M. Variable selection method for genome-wide SNP-SNP interaction analysis. 2nd International Symposium on Statistical Genetics (ISSG). Seoul; 2013 August.
 15. Ueki M. Variable selection method for whole-genome gene-gene interaction analysis. Joint Meeting of the IASC Satellite Conference for the 59th ISI WSC and the 8th Conference of the Asian Regional Section(ARS) of the IASC. Seoul; 2013 August.
- 〔図書〕(計 1 件)
1. 田宮 元, 植木優夫, 小森 理「ゲノム医学のための遺伝統計学」共立出版(2015/3 刊行)
- 〔産業財産権〕
出願状況(計 0 件)
- 名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :
- 取得状況(計 0 件)
- 名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :
- 〔その他〕
ホームページ等
- 6 . 研究組織
(1)研究代表者
植木 優夫 (UEKI, Masao)
久留米大学・バイオ統計センター・准教授
研究者番号 : 10515860
- (2)研究分担者

()

研究者番号：

(3)連携研究者

()

研究者番号：