

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 8日現在

機関番号：11501

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22590300

研究課題名（和文） ゲノムワイド遺伝子×遺伝子相互作用解析法の開発

研究課題名（英文） Methods of genome-wide gene-gene interaction analysis

研究代表者

田宮 元 (TAMIYA GEN)

山形大学・医学部・教授

研究者番号：10317745

研究成果の概要（和文）：多因子疾患の標準的な SNP-GWAS（ゲノムワイド関連解析）データを網羅的に用いて、「遺伝子×遺伝子相互作用」を検出するための、新しい遺伝統計学的手法を開発することを目的として、周辺回帰と罰則付回帰分析とを組み合わせた高次元変数選択の手法をゲノムワイド遺伝子×遺伝子相互作用解析に適用した遺伝統計アルゴリズムを開発した。また、それを安価な環境で実装し、5000人×50万 SNP の GWAS データで網羅的な相互作用検索を数時間以内に実行するソフトウェアプログラムを開発した。

研究成果の概要（英文）：We propose utilizing a model selection strategy for an exhaustive interaction analysis in whole GWA data. Our new software, EPISIS, implements the 'sure independence screening', which has two major parts, screening and variable selection; EPISIS can complete an exhaustive search for all possible SNP by SNP interactions in non-filtered GWA data within several days under common computing environments. Our fast program is applied to find novel genetic components in GWA data from WTCCC (Wellcome Trust Case-Control Consortium).

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,900,000	570,000	2,470,000
2011年度	1,000,000	300,000	1,300,000
2012年度	600,000	180,000	780,000
総計	3,500,000	1,050,000	4,550,000

研究分野：人類遺伝学

科研費の分科・細目：ゲノム科学・ゲノム医科学

キーワード：SNP-GWAS、SNPs、ゲノムワイド関連解析、モデル選択、変数選択、遺伝子×遺伝子相互作用、高次元変数選択

### 1. 研究開始当初の背景

糖尿病や精神疾患など複数の遺伝子座が寄与する多因子性疾患の多くは、それが単一遺伝子性でなければ（つまり血縁度が離れるに応じた血縁者再発リスクの急速な減衰があれば）、遺伝子座間のエピスタシス相互作用が関与するとしてモデル化されてきた。

このエピスタシス相互作用のことを「遺伝子×遺伝子相互作用」と呼ぶが、もし、個々の遺伝子が疾患発症に対して単独では有意な効果を持たず、他の遺伝子との相乗的相互作用を通してのみ強い効果を発揮するとすると、現在、世界的に主流となっている単一 SNP の検定を基にしたゲノムワイド関連解析（Genome-Wide Association Study;

GWAS) だけでは、疾患感受性遺伝子を同定できない可能性がある。

上記のような深刻な偽陰性を解決するには、単一 SNP の検定を基にするのではなく、複数の SNPs の相互作用に基づく遺伝統計解析を実施する必要がある、古典的にはロジスティック回帰分析などのパラメトリックな手法が用いられてきた。ところが SNPs や環境因子など考慮すべき因子数 ( $p$ ) が急速に増加しているにも関わらず、サンプルサイズ ( $n$ ) は増加できないという状況が生じている。古くから、 $p$  が増加すると推定精度の低下を招くことが知られており、次元の呪い (curse of dimensionality) と呼ばれていたが、昨今の SNP-GWAS などではこの状況が極端に悪化しており  $p \gg n$  問題と総称されている。この  $p \gg n$  条件下では、従来の  $p < n$  条件下で保証されていた回帰分析諸法の数学的妥当性は破綻し、その結果、検定の有意性など標準的な統計的推測はその理論的根拠を失ってしまう。このような問題の克服のために、探索的なノンパラメトリック手法である多因子次元削減やアンサンブルツリーなどの再帰的分割アプローチが開発されているが、極端な計算量が必要とされ、数十万ものゲノムワイド SNPs データでは、次世代スーパーコンピュータを用いたとしても計算実行可能性が維持できない。そこで、国内外では、一定の P 値などを閾値にして事前に SNPs 数を削減する手法が主流となっているが、これでは単独では陽性を示さない SNP が選抜に残らず、本来の相互作用が取り逃がしてしまうという深刻なジレンマに陥っていた。

## 2. 研究の目的

そこで、このジレンマを乗り越えるために、小標本高次元の SNP-GWAS データに罰則付回帰のような高次元変数選択の枠組みを適用するアプローチを着想した。罰則付回帰は、実際には、ほとんどの因子が疾患発症に対する効果を持たないこと (スパース性) を前提とした上で、罰則付きの最適化 (正則化) を行うことで、自動的にゼロ係数を生み出す回帰分析法であるこの画期的な性質によって、統計学的保証を回復しつつ、一般的な計算機資源であっても数十万の SNP-GWAS データの相互作用を妥当な計算時間内で検討することが可能になる。

## 3. 研究の方法

Lasso は標本数  $n$  を越える非ゼロ係数を抽出することは不可能であるが、誤って非ゼロとした係数がすべてゼロに近い値で収ま

るという性質も示されている。従って、少なくとも Lasso を説明変数の初期のスクリーニングとして利用可能である。変数間に強い相関がある場合には、Lasso を拡張した Elastic Net の利用も可能である。本計画では、まず上記手法や周辺回帰でスクリーニングを行い、さらにスクリーニング後に残った相互作用について、SCAD 法や Adaptive Lasso、あるいは Dantzig selector を用いることで、回帰係数を高精度に推定する。テストデータとしてもっとも頻繁に用いられている WTCCC のクローン病の SNP-GWAS データを用いる。開発されたソフトウェアを用いて、WTCCC の 7 疾患全てについて遺伝子×遺伝子相互作用を検索する。開発したソフトウェアはオープンソースとして公開する。

## 4. 研究成果

多因子疾患の標準的な SNP-GWAS (ゲノムワイド関連解析) データを網羅的に用いて、「遺伝子×遺伝子相互作用」を検出するための、新しい遺伝統計学的手法を開発することを目的として研究を行った。研究計画は、ゲノムワイド遺伝子×遺伝子相互作用解析に、1) 罰則付回帰分析を適用した遺伝統計手法を開発し、2) 一般的な計算機資源で実行可能なソフトウェアツールを開発し公開するというものであった。実際に、研究期間内に、罰則付回帰分析と周辺回帰を組み合わせた高次元変数選択の手法 (Sure Independence Screening; SIS) の手法をゲノムワイド遺伝子×遺伝子相互作用解析に適用した遺伝統計アルゴリズムを開発し、それを安価な GPGPU 環境 (General Purpose Graphic Processing Unit) で実装し、5000 人×50 万 SNP の GWAS データ (WTCCC より取得) で網羅的な相互作用検索を数時間以内に実行するソフトウェアプログラムを開発した。本研究の成果は、申請者を最終著者として BMC Bioinformatics 2012; 13(1):72 にて論文出版が行われており、特許出願も行った。また、実際に、高次元変数選択の手法を疾患データに適用した研究として、J Allergy Clin Immunol. 2012; 129(5): 1413-6 にて論文出版を行っている。以上の内容につき、国内外での複数の関連学会で発表した。

## 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 15 件)

- ① Sato H, Uchida T, Toyota K, Kanno M, Hashimoto T, Watanabe M, Nakamura T, Tamiya G, Aoki K, Hayasaka K.

- Association of breast-fed neonatal hyperbilirubinemia with UGT1A1 polymorphisms: 211G>A (G71R) mutation becomes a risk factor under inadequate feeding. *J Hum Genet.* 査読有、2013、58(1)、7-10  
DOI: 10.1038/jhg.2012.116
- ② Yamagata University Genomic Cohort Consortium. Tamiya G (51 番目)、Ueki M (52 番目)、他 57 名、Constructing a contemporary gene-environmental cohort: study design of the Yamagata Molecular Epidemiological Cohort Study. *J Hum Genet.* 査読有、2013、58(1)、54-56  
DOI: 10.1038/jhg.2012.128
- ③ Abe Y, Tamiya G, Nakamura T, Hozumi Y, Suzuki T. Association of melanogenesis genes with skin color variation among Japanese females. *J Dermatol Sci.* 査読有、2013、69(2)、167-172  
DOI: 10.1016/j.jdermsci.2012.10.016
- ④ Makino S, Masuda C, Ando S, Tamiya G, Tooyama I. Generation of a Monoclonal Antibody Specifically Reacting with Neuron-specific TATA-Box Binding Protein-Associated Factor 1 (N-TAF1). *Antibodies.* 査読有、2013、2、1-8  
DOI: 10.3390/antib2010001
- ⑤ Ueki M, Tamiya G. Ultrahigh-dimensional variable selection method for whole-genome gene-gene interaction analysis. *BMC Bioinformatics.* 査読有、2013、13、72  
DOI: 10.1186/1471-2105-13-72
- ⑥ Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. *PLoS Genet.* 査読有、2012、8(4)、e1002625  
DOI: 10.1371/journal.pgen.1002625
- ⑦ Ueta M, Tokunaga K, Sotozono C, Sawai H, Tamiya G, Inatomi T, Kinoshita S. HLA-A\*0206 with TLR3 polymorphisms exerts more than additive effects in Stevens-Johnson syndrome with severe ocular surface complications. *PLoS One.* 査読有、2012、7(8)、e43650  
DOI: 10.1371/journal.pone.0043650
- ⑧ Ueta M, Tamiya G, Tokunaga K, Sotozono C, Ueki M, Sawai H, Inatomi T, Matsuoka T, Akira S, Narumiya S, Tashiro K, Kinoshita S. Epistatic interaction between Toll-like receptor 3 (TLR3) and prostaglandin E receptor 3 (PTGER3) genes. *J Allergy Clin Immunol.* 査読有、2012、129(5)、1413-1416  
DOI: 10.1016/j.jaci.2012.01.069
- ⑨ Kato T, Tamiya G, Koyama S, Nakamura T, Makino S, Arawaka S, Kawanami T, Tooyama I. UBR5 Gene Mutation Is Associated with Familial Adult Myoclonic Epilepsy in a Japanese Family. *ISRN Neurol.* 査読有、2012、2012、508308  
DOI: 10.5402/2012/508308
- ⑩ Jambaldorj J, Makino S, Munkhbat B, Tamiya G. Sustained expression of a neuron-specific isoform of the Tafl gene in development stages and aging in mice. *Biochem Biophys Res Commun.* 査読有、2012、425(2)、273-277  
DOI: 10.1016/j.bbrc.2012.07.081
- ⑪ Ueki M, Kawasaki Y. Automatic grouping using smooth-threshold estimating equations. *Electron J Statist.* 査読有、2011、5、309-328  
DOI: 10.1214/11-EJS608
- ⑫ Takahashi Y, Iseki C, Wada M, Momma T, Ueki M, Kawanami T, Daimon M, Suzuki K, Tamiya G, Kato T. Impaired Glucose Metabolism Slows Executive Function Independent of Cerebral Ischemic Lesions in Japanese Elderly: The Takahata Study. *Intern Med.* 査読有、2011、50(16)、1671-1678  
DOI:10.2169/internalmedicine.50.4871
- ⑬ Mashima Y, Konta T, Kudo K, Suzuki K, Ikeda A, Ichikawa K, Shibata Y, Watanabe T, Tamiya G, Kato T, Kawata S, Kubota I. Polymorphism of proinflammatory cytokine genes and albuminuria in the Japanese general population: the Takahata study. *Nephrol Dial Transplant.* 査読有、2011、26(12)、3902-3907  
DOI: 10.1093/ndt/gfr105
- ⑭ 田宮元, 植木優夫, 中村智洋. ゲノムコホート研究の理論的側面、*HORMONE FRONTIER IN GYNECOLOGY.* 査読有、2011、18(1)、27-34  
[http://m-review.co.jp/magazine/detail/J17\\_18\\_01](http://m-review.co.jp/magazine/detail/J17_18_01)
- ⑮ Ueki M, Fueda K. Boosting local quasi-likelihood estimators. *Ann Inst Stat Math.* 査読有、2010、62(2)、235-248  
DOI: 10.1007/s10463-008-0173-5

[学会発表] (計 8 件)

- ① Ueki M. Genome-wide gene-gene interaction analysis based on ultrahigh-dimensional variable selection. 5th International Conference of the ERCIM WG on COMPUTING

& STATISTICS (ERCIM 2012) (招待講演)、  
2012年12月1日～2012年12月3日、オ  
ビエドカンファレンスセンター (オビエ  
ド・スペイン)

- ② Tamiya G. Molecular evolution of  
IZUMO1 and CD9; genes involved in  
reproductive fitness. International  
Symposium on the Mechanisms of Sexual  
Reproduction in Animals and Plant.  
Joint Meeting of the 2nd  
Allo-Authentication Meeting and the  
5th Egg-Coat Meeting (MCBEEC)、2012  
年11月12日～2012年11月16日、ホテ  
ル名古屋ガーデンパレス (名古屋)
- ③ 田宮元、役に立つ集団遺伝学、第5回生  
殖研究若手の会、2012年7月26日～2012  
年7月28日、東京大学三崎臨海実験所(神  
奈川)
- ④ 中村智洋、adaptive デザインのゲノム疫  
学研究への適用、日本人類遺伝学会第57  
回大会、2012年10月24日～2012年10  
月27日、京王プラザホテル (東京)
- ⑤ 植木優夫、高次元変数選択法によるゲノ  
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、日本人類遺伝学会第  
57回大会、2012年10月24日～2012年  
10月27日、京王プラザホテル (東京)
- ⑥ Kawasaki Y. Variable Selection in  
Discrete Choice Models. American  
Statistical Association. 2010年7月  
31日～2010年8月5日、カナダプレイス  
(バンクーバー・カナダ)
- ⑦ Ueki M. A machine learning approach for  
genome-wide interaction analysis. the  
60th annual meeting of the American  
Society for Human Genetics. 2010年  
11月2日～2010年11月6日、D.C. コン  
ベンションセンター (ワシントンD.C.・  
アメリカ)
- ⑧ Makino S. Neuron-specific enhanced  
expression of TAF1 and its isoform. the  
60th annual meeting of the American  
Society for Human Genetics. 2010年  
11月2日～2010年11月6日、D.C. コン  
ベンションセンター (ワシントンD.C.・  
アメリカ)

[その他]

ホームページ等

[http://www.genetix-h.com/~AMERI/index.h  
tml](http://www.genetix-h.com/~AMERI/index.html)

## 6. 研究組織

### (1) 研究代表者

田宮 元 (TAMIYA GEN)  
山形大学・医学部・教授  
研究者番号：10317745

### (2) 研究分担者

植木 優夫 (UEKI MASAO)  
山形大学・医学部・助教  
研究者番号：10515860