

令和 6 年 6 月 13 日現在

機関番号：11301

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K10439

研究課題名（和文）GWAS要約統計量に基づく高予測能疾患発症リスク予測システム

研究課題名（英文）Development of a high-performance system for disease risk prediction on GWAS summary statistics

研究代表者

成田 暁（Narita, Akira）

東北大学・東北メディカル・メガバンク機構・助教

研究者番号：50459468

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）：本研究は、ゲノムワイド関連解析（GWAS）を実施し、その結果に基づいて、疾患発症リスク予測システムの基盤を確立することを目的としている。  
GWASとは、ゲノムの型の差異が、疾患のリスクなど健康上の様々な特性（形質）に違いをもたらすかを、統計学的に探索するアプローチである。本研究では、東北メディカル・メガバンク計画（TMM）地域住民コホートのゲノムコホートデータを用いて、および他の研究機関との共同研究として、様々な形質についてGWASを実施し、その結果（要約統計量）を用いて構築されたリスクスコアに基づいて、疾患発症リスクをどの程度正確に予測できるかの検証を行った。

研究成果の学術的意義や社会的意義

本研究を通じて、国内外のゲノムコホート/バイオバンクで実施されたGWASの結果を幅広く収集し、より大規模、より多様な民族集団のデータを統合することが、疾患原因遺伝子多型の探索や疾患リスク予測の精度向上に重要であることが示された。

さらに、多様な民族集団を対象としたゲノム医学研究は、医療アクセスに関する、民族集団間の不平等な状況を是正し、次世代医療システムの恩恵を広くもたらすと考えられ、本研究はその先進的な取組みの一つと位置付けられる。

研究成果の概要（英文）：The aim of the study is to establish the basis for a high-performance disease risk prediction system, based on summary statistics of genome-wide association analysis (GWAS).

GWAS is a statistical genetic approach in which genetic variants are tested for association with a health-related trait to detect disease-causing variants. In the study, we conducted GWAS on a range of traits/diseases using the data from the Tohoku Medical Megabank Project Community-Based Cohort Study and in collaboration with other (both domestic and international) research institutions, and assessed the predictive ability of polygenic risk scores, which were calculated from the GWAS summary statistics, for various diseases.

研究分野：遺伝統計学

キーワード：ゲノムワイド関連解析

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年のゲノム解析技術の進展により、大規模かつ高密度なゲノムデータが利用可能となり、それらを用いた GWAS (genome-wide association study; 全ゲノム関連解析) などのアプローチによって、疾患を含む様々な形質と関連する SNP (single nucleotide polymorphism; 一塩基多型) などの DNA 多型が数多く同定されてきた。

それに伴い、疾患を含む様々な形質の多くは、効果の大きい少数のゲノム多型でなく、効果が微細な多数の多型によって支配されていることが示唆され、個人ごとの疾患発症リスクの新たな指標として、GWAS で得られた要約統計量から構築される PRS (polygenic risk score; ポリジェニックリスクスコア) が提唱されている。また、GWAS に用いられるデータの大規模化や祖先性 (遺伝的背景) の多様化に伴い、それぞれの GWAS の効果をメタ解析で統合し、PRS の計算に用いる戦略も定着しつつある。

### 2. 研究の目的

本研究は、これまでの、また今後もさらに蓄積される GWAS 要約統計量や、個人レベルのゲノムデータをもとに、疾患を含む各種形質の PRS に基づく疾患発症リスク予測システムの基盤を確立することを目的とする。

### 3. 研究の方法

他の研究機関との共同研究として、提供された検体のゲノム多型タイピングや imputation (遺伝子型推定) を実施した。また、TMM (東北メディカル・メガバンク計画) 地域住民コホートのゲノムコホートデータを用いて GWAS や PRS の計算を実施し、その要約統計量を共同研究機関と共有、またデータベースにて公開した。

個人レベルのゲノムデータを用いた疾患リスク予測手法については、機械学習を応用した STMGP (smooth-threshold multivariate genetic prediction; Ueki and Tamiya 2016) が提唱されている。本研究では、TMM 地域住民コホートデータおよび UK バイオバンクデータに STMGP を適用し、予測能に関して、clumping and thresholding など従来の手法との比較を行った。また、計算負荷軽減のため、データの分割が推定予測モデルに与える影響について、また、外部の GWAS 要約統計量を STMGP に取り込むことで、予測能の向上が見られるかの検証を行った。

### 4. 研究成果

共同研究で産生されたゲノムデータは、ToMMo (東北大学東北メディカル・メガバンク機構) が運営、管理するスーパーコンピュータに蓄積されており、疾患リスク予測システムの基盤確立に向けて二次利用される見通しである。

一方で、TMM 地域住民コホートデータを用いた GWAS や PRS の計算も多数実施した。Global Lipids Genetics Consortium との共同研究では、血中脂質の一種である LDL-C について、ヨーロッパ系集団を中心とする 165.5 万人のメタ GWAS から得られたゲノム多型のリストとその効果量の推定値を用いて、TMM 地域住民コホート調査 (宮城) の参加者約 28,000 人の PRS を算出し、LDL-C の分布をどの程度説明できるかを評価した。その結果、祖先性集団ごとに特定されたゲノム多型とその推定効果量よりも、全集団のデータから特定されたゲノム多型とその効果量をもとに算出された PRS の方が、LDL-C の分布をよく説明できることが示された。また、関連ゲノム多型の検出力やファインマッピングの解像度向上についても、遺伝的多様性が重要な因子であることが示された (Graham et al. 2021)。さらに、メタ GWAS の要約統計量を、染色体上の位置や機能、遺伝子発現量などのデータと組み合わせることにより、脂質レベルと生物学的に関連する遺伝子を集中的に同定できることが示された (Ramdas et al. 2022; Kanoni et al. 2022)。

この他にも、国内外の研究機関と共同して、メタボロームデータ (Sakaue et al. 2021) や腎機能関連形質 (Sugawara et al. 2022)、小児血圧、薬剤応答性、変形性膝関節症、膵がん・胆道がんなどの GWAS を実施した。これらの要約統計量は、ToMMo の jMorp データベースでも順次公開されており、その相互利用を通じて、国内外の研究機関における疾患発症リスク予測研究の推進に寄与するものと期待される。

疾患リスク予測手法については、TMM 地域住民コホートデータおよび UK バイオバンクデータを用いた検証で、STMGP が従来の手法を上回る予測能を示す結果が得られた。また、計算負荷軽減のため、データを分割し、各サブセットで最適化された予測モデルを統合しても、分割しない場合と同等の予測精度が期待できること、さらに、外部の GWAS 要約統計量を取り込むことで、予測能が向上することを見出した。

### 引用文献

Ueki M & Tamiya G. Smooth-Threshold Multivariate Genetic Prediction with Unbiased Model Selection. *Genet Epidemiol* 40:233-243, 2016.

Graham SE et al. The power of genetic diversity in genome-wide association studies of

- lipids. *Nature* 600:675-679, 2021.
- Ramdas S et al. A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. *Am J Hum Genet* 109:1366-1387, 2022.
- Kanoni S et al. Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. *Genome Biol* 23:268, 2022.
- Sakaue S et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* 53:1415-1424, 2021.
- Sugawara Y et al. Genome-wide association study of the risk of chronic kidney disease and kidney-related traits in the Japanese population: J-Kidney-Biobank. *J Hum Genet* 68:55-64, 2023.

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 2件/うちオープンアクセス 2件）

1. 著者名 Ramdas Shweta, Judd Jonathan, Graham Sarah E., Kanoni Stavroula, Wang Yuxuan, ..., Narita Akira, ..., Tamiya Gen, ..., Brown Christopher D.	4. 巻 109
2. 論文標題 A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids	5. 発行年 2022年
3. 雑誌名 The American Journal of Human Genetics	6. 最初と最後の頁 1366 ~ 1387
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ajhg.2022.06.012	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Sugawara Yuka, Hirakawa Yosuke, Nagasu Hajime, Narita Akira, Katayama Akihiro, Wada Jun, Shimizu Miho, Wada Takashi, Kitamura Hiromasa, Nakano Toshiaki, Yokoi Hideki, Yanagita Motoko, Goto Shin, Narita Ichiei, Koshiba Seizo, Tamiya Gen, Nangaku Masaomi, Yamamoto Masayuki, Kashiwara Naoki	4. 巻 68
2. 論文標題 Genome-wide association study of the risk of chronic kidney disease and kidney-related traits in the Japanese population: J-Kidney-Biobank	5. 発行年 2022年
3. 雑誌名 Journal of Human Genetics	6. 最初と最後の頁 55 ~ 64
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s10038-022-01094-1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kanoni Stavroula, Graham Sarah E., Wang Yuxuan, Surakka Ida, Ramdas Shweta, ..., Narita Akira, ..., Tamiya Gen, ..., Peloso Gina M.	4. 巻 23
2. 論文標題 Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis	5. 発行年 2022年
3. 雑誌名 Genome Biology	6. 最初と最後の頁 268
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13059-022-02837-1	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Sakaue Saori, Kanai Masahiro, Tanigawa Yosuke, Karjalainen Juha, Kurki Mitja, Koshiba Seizo, Narita Akira et al.	4. 巻 53
2. 論文標題 A cross-population atlas of genetic associations for 220 human phenotypes	5. 発行年 2021年
3. 雑誌名 Nature Genetics	6. 最初と最後の頁 1415 ~ 1424
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41588-021-00931-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Graham Sarah E., Clarke Shoa L., Wu Kuan-Han H., Kanoni Stavroula, Zajac Greg J. M., ..., Narita Akira, ..., Tamiya Gen, ..., Willer Cristen J.	4. 巻 600
2. 論文標題 The power of genetic diversity in genome-wide association studies of lipids	5. 発行年 2021年
3. 雑誌名 Nature	6. 最初と最後の頁 675 ~ 679
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41586-021-04064-3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	田宮 元  (Tamiya Gen)  (10317745)	東北大学・医学系研究科・教授    (11301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------