

令和 3 年 5 月 25 日現在

機関番号：14401

研究種目：若手研究

研究期間：2019～2020

課題番号：19K17962

研究課題名（和文）機械学習による疾患ゲノムからの2型糖尿病リスク層別化・サブタイプ分類

研究課題名（英文）Risk stratification and subtype classification of type 2 diabetes from disease genome by machine learning

研究代表者

鈴木 顕（Suzuki, Ken）

大阪大学・医学系研究科・助教

研究者番号：90838806

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：2019年度、2020年度を通じて、1．トレーニングデータを用いた2型糖尿病感受性遺伝子変異の選定、2．各種既存アルゴリズムを適用した2型糖尿病発症リスクの層別化、3．新規機械学習アルゴリズムの開発によるリスク層別化の精度向上、の3つの課題を実施した。これらの知見により、同等の2型糖尿病発症リスクを有する被検者であっても、パスウェイ・クラスターPRSに基づきその病型の評価が可能になることが期待される。

研究成果の学術的意義や社会的意義

ポリジェニック・リスク・スコア（以下、PRS）を用いた遺伝情報による疾患発症リスクの推定は欧米人集団においては進展していたが、日本人集団においては発展途上であった。本研究を通じて2型糖尿病の日本人集団におけるPRSの研究が促進され、2型糖尿病発症予測の実現に一步近づいた。また、PRSでは2型糖尿病の病型分類はできないが、本研究を通じて開発された各種アルゴリズムにより2型糖尿病の多様性を見ることが可能になった。今後2型糖尿病の病型による治療の個別化が望まれる。

研究成果の概要（英文）：Throughout FY2019 and FY2020, I have conducted three tasks: 1. selection of type 2 diabetes susceptibility gene variants using training data, 2. stratification of risk of developing type 2 diabetes by applying various existing algorithms, and 3. development of novel machine learning algorithms to improve the accuracy of risk stratification. These findings are expected to make it possible to evaluate the disease type of subjects with comparable risk of developing type 2 diabetes based on the pathway cluster PRS.

研究分野：遺伝統計学

キーワード：2型糖尿病 病型分類

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

次世代シーケンサーや高密度 SNP マイクロアレイ技術の発展と低コスト化により、数十万人～百万人規模のゲノムワイド関連解析 (genome-wide association study; GWAS) が実施されてきた。申請者らは日本人集団において大規模 GWAS を実施し、2 型糖尿病の感受性遺伝子変異をこれまでに多数同定してきた。

一方で、疾患ゲノムの臨床への活用は発展途上である。疾患ゲノムによるリスク予測やサブタイプ分類に関する先行研究は精度において臨床への応用は程遠い状況である。本研究では申請者らの疾患ゲノム研究をさらに発展させる形で、2 型糖尿病発症リスクの層別化とサブタイプ分類という問題に、斬新なアルゴリズムを適用する。

2. 研究の目的

近年、弱い効果を有する疾患感受性遺伝子変異まで考慮し、数千～数十万の遺伝子変異をモデルに含め重み付けをして総和を計算するポリジェニックリスクスコア (polygenic risk score) を用いたリスク層別化が有用であることがわかってきており、いよいよ疾患ゲノムデータによるリスク層別化が実臨床に応用されるものと、注目を集めている。

しかし、日本人集団でのポリジェニックリスクスコア解析は未実施であり、特に 2 型糖尿病は日本人と欧米人で表現型が異なるため、日本人集団でポリジェニックリスクスコア解析を実施することが極めて重要である。また、近年世界人口のごく一部しか占めない欧米人集団のみで疾患ゲノム解析のほとんどが実施されているという同質性が疾患ゲノムの理解を阻んでいるという認識が広まっており、全人口の 60% を占めるアジア人におけるポリジェニックリスクスコア解析は世界的に注目される。

本研究では、申請者らが実施してきた 2 型糖尿病疾患ゲノム研究をさらに発展させる形で、様々な斬新なアルゴリズム (高次元圧縮技術・ポリジェニックリスクスコア) を駆使し、2 型糖尿病のリスク層別化および疾患のサブタイプ分類を行う。本研究により、「遺伝情報により 2 型糖尿病の発症リスクおよびその表現型はどのように規定されるか」という疾患ゲノム研究の重要課題が明らかになることが期待される。

3. 研究の方法

本研究では「遺伝情報により 2 型糖尿病発症リスクおよびその表現型はどのように規定されるか」という医学的に重要な問いに対して、以下の 3 つの方法で取り組む。

- ：トレーニングデータを用いた 2 型糖尿病感受性遺伝子変異の選定
- ：各種既存アルゴリズムを適用した 2 型糖尿病発症リスクの層別化とサブタイプ分類
- ：新規機械学習アルゴリズムの開発によるリスク層別化・サブタイプ分類の精度向上

4. 研究成果

2019 年度には大規模ゲノム解析を実施し、2 型糖尿病に関連する遺伝子を選定した。バリエーションの選定については条件を柔軟に設定し、さまざまな選定パターンを実施した。公共データベースに登録されている 2500 人分のゲノムデータとバリエーション情報を活用し、対象集団のジェノタイプデータを推定することで、全ゲノム上のバリエーションのカバー率を向上させた。このカバー率の向上により層別化や分類の精度が格段に向上した。各種アルゴリズムによる解析を実施し、層別化や分類について新たな洞察を得た。まず、疾患の層別化については、従来より知られている疫学的な層別化に双肩しうる性能を有する可能性が明らかとなった。一方、分類についてはこれまで異質性、多様性が指摘されていたが、明らかな分類法は確立されていなかった。今年度の研究において、ゲノムデータに対して、これまで応用されることが稀であったアルゴリズム・手法を積極的に応用することにより、分類に関する新たな知見を得た。各種臨床データと各分類との関連を網羅的かつ定量的に評価し、特徴的な関連を同定した。また、オンライン上に公開されている公共データベースを用いて、分類の分子生物学的特徴を評価したところ、各分類に有意に関連する事象を同定することができ、各分類の分子生物学的特徴づけを行うことに成功した。

2020 年度は 1. トレーニングデータを用いた 2 型糖尿病感受性遺伝子変異の選定、2. 各種既存アルゴリズムを適用した 2 型糖尿病発症リスクの層別化、3. 新規機械学習アルゴリズムの開発によるリスク層別化の精度向上、の 3 つの課題を実施した。1. については、バイオバンク・ジャパン 20 万人 (うち糖尿病 4 万人、非糖尿病 16 万人) のデータを活用し、ゲノムワイド関連解析 (GWAS) を実施し、2 型糖尿病に関連する遺伝子変異を選択した。2. については、1. で同定した 2 型糖尿病感受性遺伝子変異に関するジェノタイプデータを用いてバイオバンク・ジャパンのサンプルにおいてゲノムデータの主成分 (PCA) による次元縮約やポリジェニック・リ

スク・スコア (PRS) によるリスク層別化を実施した。既存の高次元圧縮技術などのアルゴリズムが抱えている問題点を克服するため、高次元圧縮技術の組み合わせ (PCA-UMAP) を適用し、2 型糖尿病のリスク層別化の精度を向上させた。さらに、分子生物学的経路 (以下、パスウェイ) のデータベースからパスウェイ情報を網羅的に取得し、各パスウェイに属する遺伝子が存在するゲノム領域の SNP に限定してポリジェニック・リスク・スコアを計算した (パスウェイ PRS)。パスウェイ PRS 同士の相関係数を算出し、相関が強いパスウェイ同士を同一のクラスターに所属させるクラスタリングを実施し、パスウェイのクラスターごとに PRS を算出した。パスウェイ・クラスター PRS を代謝疾患に関わる形質との関連に基づき特徴付を行った。パスウェイ・クラスター PRS の算出により、同一の PRS を有する被検者であっても、そのパスウェイ・クラスター PRS の分布は異なることが明らかとなった。上記結果は、2 型糖尿病の病態の理解を促進し、2 型糖尿病の診療改善に寄与することが期待される。

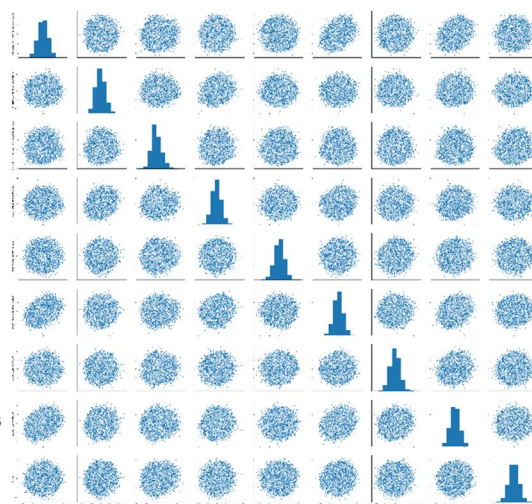
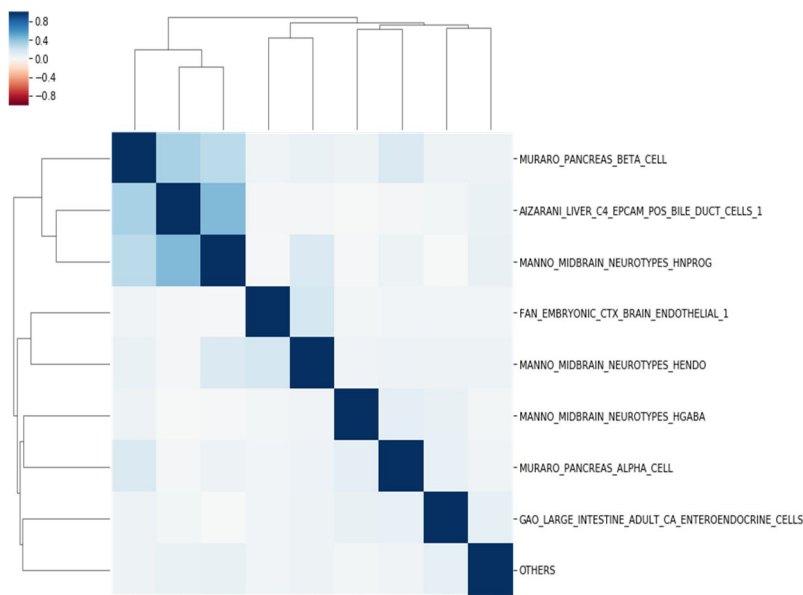


図 1 2 型糖尿病 パスウェイ PRS の相関係数のヒートマップ

図 2 2 型糖尿病 パスウェイ PRS のペアプロット

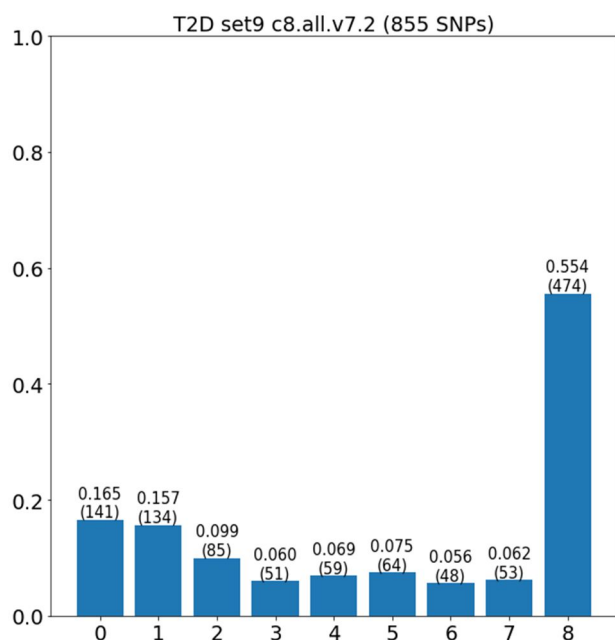


図 3 各パスウェイの 2 型糖尿病 PRS SNP のカバー率

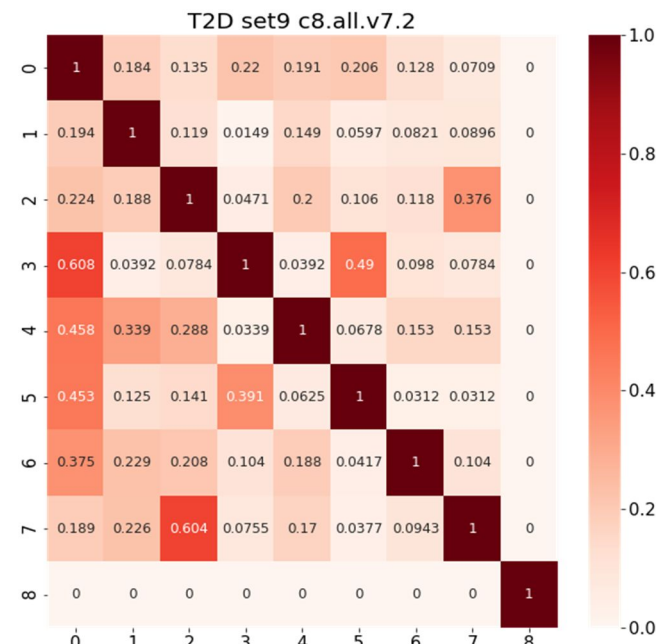


図 4 各パスウェイ同士の 2 型糖尿病 PRS SNP の共有率

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 12件／うち国際共著 6件／うちオープンアクセス 3件）

1. 著者名 Yasumizu, Yoshiaki; Sakaue, Saori; Konuma, Takahiro; Suzuki, Ken; Matsuda, Koichi; Murakami, Yoshinori; Kubo, Michiaki; Palamara, Pier Francesco; Kamatani, Yoichiro; Okada, Yukinori	4. 巻 37
2. 論文標題 Genome-Wide Natural Selection Signatures Are Linked to Genetic Risk of Modern Phenotypes in the Japanese Population	5. 発行年 2020年
3. 雑誌名 Molecular Biology and Evolution	6. 最初と最後の頁 1306-1316
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/molbev/msaa005	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 該当する
1. 著者名 Sakaue, Saori; Suzuki, Ken; et al.	4. 巻 11
2. 論文標題 Dimensionality reduction reveals fine-scale structure in the Japanese population with consequences for polygenic risk prediction	5. 発行年 2020年
3. 雑誌名 Nature Communications	6. 最初と最後の頁 1569
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41467-020-15194-z	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 Lin, Yingsong; Suzuki, Ken; et al.	4. 巻 11
2. 論文標題 Genome-wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer	5. 発行年 2020年
3. 雑誌名 Nature Communications	6. 最初と最後の頁 3175
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41467-020-16711-w	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 -
1. 著者名 Ishigaki, Kazuyoshi; Suzuki, Ken; et al.	4. 巻 52
2. 論文標題 Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases	5. 発行年 2020年
3. 雑誌名 Nature Genetics	6. 最初と最後の頁 669-679
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41588-020-0640-3	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 -

1. 著者名 Spracklen, Cassandra N.;Suzuki, Ken; et al.	4. 巻 582
2. 論文標題 Identification of type 2 diabetes loci in 433,540 East Asian individuals	5. 発行年 2020年
3. 雑誌名 Nature	6. 最初と最後の頁 240-245
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41586-020-2263-3	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 該当する

1. 著者名 Hosoe, Jun; Miya,Suzuki, Ken; et al.	4. 巻 169
2. 論文標題 Clinical usefulness of multigene screening with phenotype-driven bioinformatics analysis for the diagnosis of patients with monogenic diabetes or severe insulin resistance	5. 発行年 2020年
3. 雑誌名 Diabetes Research and Clinical Practice	6. 最初と最後の頁 108461
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.diabres.2020.108461	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 Kishikawa, Toshihiro; Suzuki, Ken; et al.	4. 巻 33
2. 論文標題 Increased levels of plasma nucleotides in patients with rheumatoid arthritis	5. 発行年 2020年
3. 雑誌名 International Immunology	6. 最初と最後の頁 119-124
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/intimm/dxaa059	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 Ogawa, Kotaro; Okuno, Tatsusada; Hosomichi, Kazuyoshi; Hosokawa, Akiko; Hirata, Jun; Suzuki, Ken; Sakaue, Saori; Kinoshita, Makoto; Asano, Yoshihiro; Miyamoto, Katsuichi	4. 巻 16
2. 論文標題 Next-generation sequencing identifies contribution of both class I and II HLA genes on susceptibility of multiple sclerosis in Japanese	5. 発行年 2019年
3. 雑誌名 Journal of neuroinflammation	6. 最初と最後の頁 162
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12974-019-1551-z	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 -

1. 著者名 Ogawa, Kotaro; Stuart, Philip E; Tsoi, Lam C; Suzuki, Ken; Nair, Rajan P; Mochizuki, Hideki; Elder, James T; Okada, Yukinori	4. 巻 139
2. 論文標題 A Transethnic Mendelian Randomization Study Identifies Causality of Obesity on Risk of Psoriasis	5. 発行年 2019年
3. 雑誌名 The Journal of investigative dermatology	6. 最初と最後の頁 1397-1400
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jid.2018.11.023	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 該当する

1. 著者名 Kazuyoshi Ishigaki, Ken Suzuki, et al.	4. 巻 52
2. 論文標題 Large scale genome-wide association study in a Japanese population identified 45 novel susceptibility loci for 22 diseases	5. 発行年 2020年
3. 雑誌名 Nature Genetics	6. 最初と最後の頁 669-679
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41588-020-0640-3	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 該当する

1. 著者名 Cassandra N. Spracklen, Ken Suzuki, et al.	4. 巻 582
2. 論文標題 Identification of type 2 diabetes loci in 433,540 East Asian individuals	5. 発行年 2020年
3. 雑誌名 Nature	6. 最初と最後の頁 240-245
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41586-020-2263-3	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 該当する

1. 著者名 Yoshiaki Yasumizu, Saori Sakaue, Takahiro Konuma, Ken Suzuki, Koichi Matsuda, Yoshinori Murakami, Michiaki Kubo, Pier Francesco Palamara, Yoichiro Kamatani, Yukinori Okada	4. 巻 37
2. 論文標題 Natural Selection Signatures Are Linked to Genetic Risk of Modern Phenotypes in the Japanese Population	5. 発行年 2020年
3. 雑誌名 Molecular Biology and Evolution	6. 最初と最後の頁 1306-1316
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/molbev/msaa005	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
--	---------------------------	-----------------------	----

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
---------	---------