

令和 2 年 6 月 2 日現在

機関番号：11301

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K12772

研究課題名(和文)ゲノム環境ワイド関連解析GE-WASによる遺伝子環境相互作用の同定

研究課題名(英文)Identification of gene-environment interactions by genome-environment wide association analysis GE-WAS

研究代表者

永家 聖(Nagaie, Satoshi)

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：00726466

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：2型糖尿病における疾患関連遺伝因子と疾患関連環境因子の複雑な相互作用であるゲノム環境ワイド関連解析GE-WAS(Gene-Environment-Wide-Association-Study)を、Cochran-Mantel-Haenszelカイニ乗検定にて行い、得られたP値でGxEランドスケーププロットを描画し、視覚的に3因子の関係性を捉えた。2型糖尿病の病型分類に必要な検体検査値(グルコース、HbA1cなど)、投薬情報(PMDAにて承認された処方薬)、レセプト病名(ICD10コード:E10-E14)を用いて、ルールベースのアルゴリズムを作成し、表現型とすることで精度を高めた。

研究成果の学術的意義や社会的意義

2型糖尿病は、遺伝因子と環境因子の複雑な相互作用により発症する疾患である。その関連性を把握することはとても重要であり、今回作成したGxEランドスケーププロットを用いることで、遺伝因子、環境因子、2型糖尿病の関係性を簡便に把握することができるものである。

また、今回2型糖尿病の有無について、治療のために用いられる診断病名ではなく、検体検査値や投薬情報等を用いた客観的な病型分類を行うことで、より精度の高いGxEランドスケーププロットを作成した。

研究成果の概要(英文)：A complex interaction between disease-related genetic factors and disease-related environmental factors in type 2 diabetes, the Gene-Environment-Wide-Association Study (GE-WAS), was performed with the Cochran-Mantel-Haenszel chi-square test to draw a GxE landscape plot with the obtained P values and visually capture the relationship between the three factors.

A rule-based algorithm was created and phenotyped using laboratory test values (e.g., glucose, HbA1c), medication information (PMDA-approved prescription drugs), and disease names (ICD10 code: E10-E14) for type 2 diabetes mellitus.

研究分野：生命情報

キーワード：フェノタイピング 遺伝子環境相互作用 医療情報 生命情報 コホート

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

家系調査による連鎖解析によって見出された希少メンデル型遺伝疾患は、オッズ比は非常に高いが頻度が非常に低い。またゲノムワイド関連解析(GWAS)で同定された疾患は、頻度は 1-5%と連鎖解析よりも高いが、オッズ比は非常に低いことが、これまでに国内外の研究者により多数報告されている。しかしながらその間となる変異アレル・遺伝子機能に与える影響が中程度の疾患については、未だ十分に解明されておらず、この問題を解決するためには遺伝子環境相互作用を考慮することが重要であると考えられている。例えば、糖尿病などのありふれた疾患は、連鎖解析やゲノムワイド関連解析を用いることでその素因遺伝子が判明されてきたが、その原因を完全に説明するまでには至っていない。

### 2. 研究の目的

疾患関連遺伝因子と疾患関連環境因子が相加的ではなく、組み合わせ特異的に疾患発症に寄与していることを確かめるために、その関係性を網羅的に可視化した三次元 GxE ランドスケーププロットを作成することで、新たな疾患関連因子を同定する。

### 3. 研究の方法

#### (1) 三次元 GxE ランドスケーププロットの作成

ゲノム情報や健康調査情報(調査票データ、検体検査データ)から、遺伝因子・環境因子・疾患情報を抽出し、三次元 GxE ランドスケーププロットを作成し、組み合わせ特異的な項目を抽出する。

ゲノム情報 : 全ゲノム情報、SNP アレイ情報  
健康調査情報 : 調査票(生活・食)情報、検体検査情報  
疾患情報 : 調査票(生活)情報の既往歴  
統計学的検定 : Cochran-Mantel-Haenszel カイ二乗検定  
次元圧縮方法 : t-SNE、UMAP、PCA  
プログラム : R

#### (2) 2型糖尿病の病型分類アルゴリズムの作成

疾患情報について、調査票における本人申告の既往歴ではなく、客観的な情報である検体検査値や投薬情報等を用いた病型分類を行うことで、より精度の高い GxE ランドスケーププロットを作成する。

#### 2型糖尿病の病型分類アルゴリズムの作成:

データセット : 疫学レセプトデータベースである JMDC Claims Database を用いる。  
このデータセットには、複数の健康保険組合より寄せられたレセプト(入院、外来、調剤)および健診データが含まれている。対象データの期間は、2005年1月から2019年5月に登録されたデータである。

#### 使用する情報

検体検査値 : グルコース、グリコアルブミン、HbA1c  
投薬情報 : PMDA にて承認された処方薬のうち、添付文書の効能・効果欄に「糖尿病」と記載がある薬を用いる。

レセプト病名 : ICD10 コード E10-E14

#### 病型分類の手法

ルールベース

評価方法 : 感度・特異度・陽性適中率を算出し、アルゴリズムの性能を評価する。

### 4. 研究成果

#### (1) 三次元 GxE ランドスケーププロットの作成

・GxE ランドスケーププロットの改良  
パイロットで作成した GxE ランドスケーププロットは、すべての P 値をそのまま利用しているため、ランドスケープとしては滑らかになっていない。そこである閾値よりも P 値が高いデータに対しては、一定値に置き換えることで、より可読性の高い GxE ランドスケーププロットを作成した。

#### (2) 2型糖尿病の病型分類アルゴリズムの作成

・処方薬  
添付文書の効能・効果欄に「糖尿病」と記載がある薬は、一般名として 89 種類の薬品が承認済みであった。

- ・疾病群

今回使用したデータは、匿名化済みのためカルテ情報を用いた医師によるバリデーションができない。そのため、病型分類の予測モデルを確率的ゴールドスタンダードにより定義した。

作成したプログラムや ReadMe 等を下記 github 上で公開し、継続的に更新を行う予定である。

<https://github.com/snagaie/gewas>

<https://github.com/snagaie/phenotyping>

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Satoshi Nagaie, Soichi Ogishima, and Hiroshi Tanaka
2. 発表標題 A method to associate all possible combinations of genetic and environmental factors using GxE landscape plot
3. 学会等名 AROB Workshop (国際学会)
4. 発表年 2017年

〔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
--	---------------------------	-----------------------	----