

令和元年6月5日現在

機関番号：11301

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K00295

研究課題名(和文) 前向きコホート、医療情報、個人の疾患感受性多型情報を統合するオントロジーの開発

研究課題名(英文) Development of ontological representation for disease and cohort information

研究代表者

高井 貴子(五十嵐貴子)(Takai, Takako)

東北大学・東北メディカル・メガバンク機構・非常勤講師

研究者番号：60222840

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：個別化予防を目指したゲノムコホート研究で収取されるデータの有効な利活用を目的として、環境因子と遺伝因子の相互作用から発症に至るプロセスおよび因果関係の概念化を行った。Basic Formal Ontologyを核としたaxiom-basedのオントロジーと、環境因子と遺伝因子の疾患に対する役割(ロール)について考察したオントロジーを構築した。ロール概念の考察により、コホートで収集する多種類のデータの間の暗黙的な生物学的関係を明示化することができた。

研究成果の学術的意義や社会的意義

オントロジーは、人間が共有する背景知識を計算機で利用できるよう、演算処理を可能としたものである。個人のゲノム配列情報を利用して個々人の生得の体質に基づく医療の提供を目指すゲノム医療の実現には、疾患、症状、検査、医薬品、生活習慣、栄養、社会活動、等の人間生活を取り巻く様々なデータを解釈するための背景知識を計算機で利用可能とするオントロジーの開発が必要とされている。

研究成果の概要(英文)：An ontology was developed to provide explicit definitions of biomedical concepts obtained in the research on the personalized genome medicine. The concepts of processes and causal relationships are mostly investigated in this study. An ontology based on the axiom theory was firstly constructed in an application of Basic Formal Ontology. Another ontology based on the rule-concept theory was developed in order to specify the roles of individual concepts in their causal relationships to the disease onset and progression. The role-concept theory has an advantage in computation of causal mechanisms among biomedical concepts observed in genome based cohort studies.

研究分野：データサイエンス

キーワード：オントロジー ゲノムコホート 個別化予防

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

医学/臨床科学の領域では、医療情報の利用を目的としたオントロジーの必要性について古くから認識されており、ICD(1948~)、SNOMED CT(1966~)、UMLS(1986~)に代表されるオントロジーが開発され利用されてきた。この流れにおける顕著な進歩は、BFO (Basic Formal Ontology) および IAO (Information Artifact Ontology) の開発である。BFO は、生命医科学領域の概念を定義するための、トップオントロジーである。IAO は、生命医学領域の特徴である、情報とデータ等の表現の問題を定義するオントロジーである。表現物と事態を峻別して、両者の関係性を明確化する道筋を与えている。さらに、医療を対象とした OGMS (Ontology for General Medical Science) が開発され、疾患 disease の本格的な定義が与えられるとともに、診断、臨床知見、症状、疾患遺伝子、発症リスク、疾患原因、病態発生、病態進行、等の医療の主要概念が、BFO と IAO の枠組みの中で、定義された。

一方、コホート、中でも健常人をリクルートして疾患の発症を観測する前向きコホートについては、オントロジー開発がなされていない領域として残されている。東北メディカル・メガバンク事業は、世界でも有数の大規模な前向きコホートを基盤とし、個人ゲノム解析機能を内包する複合バイオバンクを開発している。本活動において収集する、生活環境の情報と個人ゲノム情報、および追跡において発見される疾患の発症、治療とその結果、予後、疾患感受性多型、疾患の浸透率、疾患リスク計算、等について、明確な概念化を行い、オントロジーを構築することは、本事業で収集する情報の有機的な結合を可能とし、すべての情報を横断する意味検索を可能として、収集した情報を有効に活用する基盤を作るものである。

### 2. 研究の目的

東北メディカル・メガバンク事業では、15万人の健常人を対象として、生活環境の情報、医療情報、個人ゲノム情報の統合データベースを構築する。前向きコホートである特徴から、未病における生活環境の情報とその推移、発症・治療・予後に関する医療情報が豊富に収集できる、ゲノム解析機能を内包することから個人ゲノムと疾患感受性多型および発症リスクの情報を豊富に収集できる。本研究では、これらの収集情報の明確な概念化を行い、オントロジーを構築することにより、本事業で収集する情報の有機的な結合を可能とし、収集した情報を有効に活用する基盤作りを目指す。

### 3. 研究の方法

東北メディカル・メガバンク事業で収集する、未病における生活環境の情報とその推移、発症・治療・予後に関する医療情報、および個人ゲノム情報と疾患感受性多型情報について、BFO/IAO/OGMS の枠組みの下、オントロジーを開発する。

### 4. 研究成果

#### (1)BFOを核とした静的概念の体系化

東北メディカル・メガバンク事業で収集する、未病における生活環境の情報とその推移、発症・治療・予後に関する医療情報、および個人ゲノム情報と疾患感受性多型情報について、BFO/IAO/OGMS の枠組みの下、オントロジーを開発した。BFO (Basic Formal Ontology) および IAO (Information Artifact Ontology) をトップオントロジーとし、既存のオントロジーで定義済みの概念については、その再利用を基本とした。BFO, RO, IAO, PATO, OGMS, OGSF, MP, HP, GO-BP, UBERON, OMIABIS より概念を取得して再構成した。続いて、被災が発症原因となる PTSD (Post Traumatic Stress Disorder) および緑内障、歯周炎を対象として、未病から発症に至るプロセスの概念化を行った。疾患の表現モデルは、Smith および Mizoguchi の定義を基本とし、コホート研究で収集する質問票や検査データ、および生体分子の機能との関係を追加した。Disease は Disease Course により具現され、Disease は Abnormal State から構成され、Disease Course は Biological Process から構成されるモデルを考えた。続いて PTSD、緑内障、歯周炎を対象として、Abnormal State である概念クラス、および Biological Process である概念クラスについて、既存のオントロジーである MP および GO-BP に対してマッピングを行い、ほぼ全ての概念クラスがマップされ得るという結果を得た。この結果は、本モデルの妥当性・有用性を示すと考える。MP および GO-BP は遺伝子のアノテーションが充実しており、遺伝子の実験データ解析に頻用される。本研究により、遺伝子の解析データから MP および GO-BP へのマッピングを介して、関連する疾患を推論できる新しい枠組みの可能性を示すことができた。

#### (2)ロール概念を核とした動的概念の体系化

個別化予防を対象とした概念整理においては、環境因子と遺伝因子の相互作用から発症に至る「プロセスおよび因果関係」の概念化が必要となる。この概念化については、OBO 上位オントロジーモデル (axiom ベースモデル) では、意味のあるプロセスのインスタンス生成ができないことに気付き、ロール概念を深く考察したフレームベースの表現が必要と判断した。ゲノムコホート調査データは、プロセスの「観測」であると位置づけ、その概念世界は function と disorder の概念階層と各階層概念の観測手段と観測結果 (データ) であるとモデル化し、各概念のロール関係を明確化しオントロジーを構築した。オントロジーエディタは法造を利用した。再構築したオントロジーでは、発症に関連する機能を定義し、それ機能の正常/異常の程度を測定・判断す

るための手段、根拠、原理を定義し、それをロール関係の考察に基づき体系化した。この体系化により、異なる手段で測定されたデータが「発症に関連する機能」ごとにグループ化され、これまで暗黙的であった因果関係を明示化することができた。例えば、高コレステロール血症のロール概念として、LDLR や PCSK9 等の遺伝型データ、栄養摂取や運動習慣の生活習慣データ、腹囲や BMI 等の生理機能検査データ、総コレステロール値やトリグリセリド値等の血液検査データ、等が高コレステロール血症の発症において担う役割を明示化した。

本研究で当初に設定した問いについては、以下に結論した。ア) アンケートは、医療領域の表現と連続した形式化が可能である。イ) アンケート(表現) 対象者の表現型(質) および個人ゲノムの多型(もの)の関係はロール概念を用いて形式化可能である。ウ) 未病から発症に至るプロセスの定義には、axiom ベースの OBO 表現に加えてロール概念が必要となる。エ) 未病から発症に至る過程と、疾患感受性多型および発症リスクの関係性について明確化するには、ファクトが不足している状況である。

## 5 . 主な発表論文等

{ 雑誌論文 }( 計 11 件 )

- [1] Sakurai R, Ueki M, Makino S, Hozawa A, Kuriyama S, Takai-Igarashi T, Kinoshita K2., Yamamoto M, Tamiya G. **Outlier detection for questionnaire data in biobanks.** *Int J Epidemiol.* 2019, dyz012. 査読有. doi: 10.1093/ije/dyz012
- [2] Miura E, Tsuchiya N, Igarashi Y, Arakawa R, Nikkuni E, Tamai T, Tabata M, Ohkouchi S, Irokawa T, Ogawa H, Takai-Igarashi T, Suzuki Y, Kuriyama S, Tamiya G, Hozawa A, Yamamoto M, Kurosawa H. **Respiratory resistance among adults in a population-based cohort study in Northern Japan.** *Respir Investig.* 2019, **57**, 274. 査読有. doi: 10.1016/j.resinv.2018.12.008.
- [3] Sugawara J, Ochi D, Yamashita R, Yamauchi T, Saigusa D, Wagata M, Obara T, Ishikuro M, Tsunemoto Y, Harada Y, Shibata T, Mimori T, Kawashima J, Katsuoka F, Takai-Igarashi T, Ogishima S, Metoki H, Hashizume H, Fuse N, Minegishi N, Koshihara S, Tanabe O, Kuriyama S, Kinoshita K, Kure S, Yaegashi N, Yamamoto M, Hiyama S, Nagasaki M. **Maternity Log study: a longitudinal lifelog monitoring and multiomics analysis for the early prediction of complicated pregnancy.** *BMJ Open.* 2019, **9**, e025939. 査読有. doi: 10.1136/bmjopen-2018-025939.
- [4] Koshihara S, Motoike I, Saigusa D, Inoue J, Shirota M, Katoh Y, Katsuoka F, Danjoh I, Hozawa A, Kuriyama S, Minegishi N, Nagasaki M, Takai-Igarashi T, Ogishima S, Fuse N, Kure S, Tamiya G, Tanabe O, Yasuda J, Kinoshita K, Yamamoto M. **Omics research project on prospective cohort studies from the Tohoku Medical Megabank Project.** 2018, *Genes Cells.* **23**, 406. 査読有. doi: 10.1111/gtc.12588.
- [5] Tanikawa C, Kamatani Y, Toyoshima O, Sakamoto H, Ito H, Takahashi A, Momozawa Y, Hirata M, Fuse N, Takai-Igarashi T, Shimizu A, Sasaki M, Yamaji T, Sawada N, Iwasaki M, Tsugane S, Naito M, Hishida A, Wakai K, Furusyo N, Murakami Y, Nakamura Y, Imoto I, Inazawa J, Oze I, Sato N, Tanioka F, Sugimura H, Hirose H, Yoshida T, Matsuo K, Kubo M, Matsuda K. **Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21.** *Cancer Sci.* 2018, **109**, 4015. 査読有. doi: 10.1111/cas.13815.
- [6] Yasuda J, Kinoshita K, Katsuoka F, Danjoh I, Sakurai-Yageta M, Motoike IN, Kuroki Y, Saito S, Kojima K, Shirota M, Saigusa D, Otsuki A, Kawashima J, Yamaguchi-Kabata Y, Tadaka S, Aoki Y, Mimori T, Kumada K, Inoue J, Makino S, Kuriki M, Fuse N, Koshihara S, Tanabe O, Nagasaki M, Tamiya G, Shimizu R, Takai-Igarashi T, Ogishima S, Hozawa A, Kuriyama S, Sugaw

ara J, Tsuboi A, Kiyomoto H, Ishii T, Tomita H, Minegishi N, Suzuki Y, Suzuki K, Kawame H, Tanaka H, Taki Y, Yaegashi N, Kure S, Nagami F; Tohoku Medical Megabank Project Study Group, Kosaki K, Sutoh Y, Hachiya T, Shimizu A, Sasaki M, Yamamoto M. **Genome analyses for the Tohoku Medical Megabank Project toward establishment of personalized healthcare.** *J Biochem.* 2018, **165** (2). 査読有. doi: 10.1093/jb/mvy096.

[7] Y Yamaguchi-Kabata, J Yasuda, O Tanabe, Y Suzuki, H Kawame, N Fuse, M Nagasaki, Y Kawai, K Kojima, F Katsuoka, S Saito, I Danjoh, I Motoike, R Yamashita, S Koshiba, D Saigusa, G Tamiya, S Kure, N Yaegashi, Y Kawaguchi, F Nagami, S Kuriyama, J Sugawara, N Minegishi, A Hozawa, S Ogishima, H Kiyomoto, T Takai-Igarashi, ToMMo Study Group, K Kinoshita, and M Yamamoto. **Evaluation of reported pathogenic variants and their frequencies in a Japanese population based on a whole-genome reference panel of 2,049 individuals.** *Journal of Human Genetics.* 2018, **63**, 213. 査読有. doi: 10.1038/s10038-017-0347-1.

[8] X Yuan, K Tsujimoto, K Hashimoto, K Kawahori, N Hanzawa, M Hamaguchi, T Seki, M Nawa, T Ehara, Y Kitamura, I Hatada, M Konishi, N Itoh, Y Nakagawa, H Shimano, T Takai-Igarashi, Y Kamei, Y Ogawa. **Epigenetic modulation of Fgf21 in the perinatal mouse liver ameliorates diet-induced obesity in adulthood.** *Nature Communications.* 2018, **9**:636. 査読有. doi: 10.1038/s41467-018-03038-w.

[9] Takai-Igarashi T, Kinoshita K, Nagasaki M, Ogishima S, Nakamura N, Nagase S, Nagaie S, Saito T, Nagami F, Minegishi N, Suzuki Y, Suzuki K, Hashizume H, Kuriyama S, Hozawa A, Yaegashi N, Kure S, Tamiya G, Kawaguchi Y, Tanaka H, Yamamoto M. **Security controls in an integrated Biobank to protect privacy in data sharing: rationale and study design.** *BMC Medical Inform Decis Mak.* 2017, **17**, 100: 査読有. doi: 10.1186/s12911-017-0494-5.

[10] S Koshiba, I Motoike, K Kojima, T Hasegawa, M Shirota, T Saito, D Saigusa, I Danjoh, F Katsuoka, S Ogishima, Y Kawai, Y Yamaguchi-Kabata, M Sakurai, S Hirano, J Nakata, H Motohashi, A Hozawa, S Kuriyama, N Minegishi, M Nagasaki, T Takai-Igarashi, N Fuse, H Kiyomoto, J Sugawara, Y Suzuki, S Kure, N Yaegashi, O Tanabe, K Kinoshita, J Yasuda & M Yamamoto. **The structural origin of metabolic quantitative diversity.** *Scientific Reports* 2016 **6**,31463. 査読有. doi: 10.1038/srep31463.

[11] S Kuriyama, N Yaegashi, F Nagami, T Arai, Y Kawaguchi, N Osumi, M Sakaida, Y Suzuki, K Nakayama, H Hashizume, G Tamiya, H Kawame, K Suzuki, A Hozawa, N Nakaya, M Kikuya, H Metoki, I Tsuji, N Fuse, H Kiyomoto, J Sugawara, A Tsuboi, S Egawa, K Ito, K Chida, T Ishii, H Tomita, Y Taki, N Minegishi, N Ishii, J Yasuda, K Igarashi, R Shimizu, M Nagasaki, S Koshiba, K Kinoshita, S Ogishima, T Takai-Igarashi, T Tominaga, O Tanabe, N Ohuchi, T Shimosegawa, S Kure, H Tanaka, S Ito, J Hitomi, K Tanno, M Nakamura, K Ogasawara, S Kobayashi, K Sakata, M Satoh, A Shimizu, M Sasaki, R Endo, K Sobue, and M Yamamoto. **The Tohoku Medical Megabank Project: Design and Mission.** *J Epidemiol* 2016, **26**,493-511. 査読有. doi: 10.2188/jea.JE20150268.

〔学会発表〕(計 20 件)

- [1] 高井貴子. Genome Medical Projects in Japan. 生物試料の保管及び活用に関する検討会. 広島. 2019.
- [2] T Takai-Igarashi. Tohoku Medical Megabank Project: a project for personalized preventive medicine based on a biobank and prospective genome cohort studies. Workshop on Bioinformatics of Gene Regulation. Goettingen. Germany 2018.
- [3] T Takai-Igarashi. GEM-J Project. Bio-Hackathon2018. Izumo. 2018.
- [4] 高井貴子. Ontology in Bio-medicine. バイオにおけるビッグデータ. 奈良. 2018.
- [5] 高井貴子. ゲノム医療研究プロジェクト. データ関連人材プログラム. 大阪. 2018.
- [6] 高井貴子. ゲノム医療研究を推進するオントロジー研究. 医療情報分析の実際. 東京. 2018.
- [7] 高井貴子. 医療オントロジー利用の現状と展望. 第 10 回 Pharma AI Forum(PHAIFO). 東京. 2017.
- [8] T Takai-Igarashi, K Kinoshita, M Nagasaki, S Ogishima, N Nakamura, S Nagase, S Nagaie, T Saito, F Nagami, N Minegishi, Y Suzuki, K Suzuki, H Hashizume, S Kuriyama, A Hozawa, N Yaegashi, S Kure, G Tamiya, Y Kawaguchi, H Tanaka, and M Yamamoto. Security Controls in an Integrated Biobank to Protect Privacy in Data Sharing. International Workshop on Artificial Life and Robotics. Korea. 2017.
- [9] T Takai-Igarashi. Aim and Progress of Tohoku Medical Megabank Project. Environmental Hazards in Multigenerational Disease -Epigenetic Mechanisms Strategic Planning Conference. Utah. USA 2017.

〔図書〕(計 3 件)

- [1]高井貴子. ゲノムオントロジー. 人工知能学大事典. pp982-983. 2017.
- [2]高井貴子. 個別化医療・個別化予防を推進するオントロジー..CICSJ Bulletin. 35. p.174-179. 2017.
- [3]高井貴子. バイオバンクとセキュリティ. 疾患データベースとバイオバンク. 実験医学増刊号. 2017.

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等なし

## 6 . 研究組織

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。