

次世代バイオデータ基盤の構築に向けた横断検索システム・データカタログ改修

コホート拡充のための大規模ゲノムコホートのデータ項目の標準化とユーザビリティ向上

研究分担者	荻島 創一	東北大学東北メディカル・メガバンク機構医療情報 ICT 部門	教授
	寶澤 篤	東北大学東北メディカル・メガバンク機構予防医学・疫学部門	教授
	中谷 直樹	東北大学東北メディカル・メガバンク機構予防医学・疫学部門	教授

研究要旨

次世代バイオデータ基盤の構築に向けて、令和元年度に実施した調査研究で取りまとめた「コホート・バイオバンクの産業利活用促進策」の具体化のために、コホート・バイオバンクの産業利用促進のための調査研究を踏まえてこれまで本研究で研究開発したコホート横断検索システムカタログを拡充するため、わが国の主要なコホート研究である5箇所のゲノムコホート拠点の協力を得て、各コホートが収集しているデータ項目とそのメタデータを標準化した。また、コホート横断検索システムカタログの機能を拡充し、ユーザビリティを向上した。

A. 研究目的

バイオ戦略では、「医療とヘルスケアが連携した末永く社会参加できる社会」の実現を目指し、「大規模統合コホート・バイオバンクの構築」のため、「健常人コホート等の実施主体が連携し、データを統合・強化する大規模健常人コホート・バイオバンクの構築」、すなわち、次世代バイオデータ基盤を構築することとされた。

このような政策的位置付けの中で、次世代バイオデータ基盤の構築に向けて、令和元年度に実施した調査研究で取りまとめた「コホート・バイオバンクの産業利活用促進策」をとりまとめた。そのなかで、産業界からのコホートの利活用にあたってのハードルの一つとして、わが国の何処のコホートにいかなるデータがあるかわからないという課題が指摘された。そこで、コホートと収集しているデータ項目とそのメタデータを検索するコホート横断検索システムカタログを昨年度、研究開発し、元施策である官民研究開発投資拡大プログラム(PRISM) バイ

オ技術領域の「糖尿病個別化予防を加速するマイクロバイーム解析 AI の開発」、「認知症に関与するマイクロバイーム・バイオマーカー解析」、さらに東北メディカル・メガバンク計画の3拠点のコホートのカタログを公開した。今年度は、コホート横断検索システムのカタログを拡充するため、産業界からの要望が強いゲノムコホートから5拠点のコホート拠点の協力を得て、各コホートから収集しているデータ項目とそのメタデータを標準化することを目的とした。また、コホート横断検索システムカタログの機能を拡充し、ユーザビリティを向上することを目的とした。

B. 研究方法

1. コホート横断検索システムカタログの拡充のための大規模ゲノムコホートのデータ項目とメタデータの標準化

コホートと収集しているデータ項目のメタデータ

を検索するコホート横断検索システムカタログを3拠点から8拠点に拡充するため、産業界からの要望が強い5拠点のコホートの協力を得て、各コホートから収集しているデータ項目とそのメタデータの標準化を行う。標準化にあたっては、昨年度実施した、官民研究開発投資拡大プログラム(PRISM) バイオ技術領域「糖尿病個別化予防を加速するマイクロバイーム解析 AI の開発」、「認知症に關与するマイクロバイーム・バイオマーカー解析」と標準化した、データ項目とそのメタデータの標準に準拠する。

今回、協力を要請するのは、産業界からの要望が強いゲノムコホートから下記の5拠点である。

- (1) 多目的コホート(JPHC)研究
- (2) 日本多施設共同コホート研究(J-MICC)研究
- (3) 愛知県がんセンター病院疫学(HERPACC)研究
- (4) 神奈川県みらい未病コホート研究
- (5) 慶應・鶴岡メタボロームコホート(TMC)研究

いずれも 10,000 人規模のコホート研究を実施し、データを収集または収集中のコホートである。

各コホート拠点には

- ◇ 委託先機関が保有する対象コホートデータの整備
- ◇ 対象コホートについて、産業界を含めた外部に対し所定の手続きのうえで提供可能なデータ項目とそのメタデータおよび性・年齢階級別人数の所定の形式での提供
- ◇ 対象コホートについての説明文書の作成
- ◇ 対象コホートに関して産業界からの照会への対応方法及び委託先機関における窓口の情報の提供

を依頼して、各コホートから収集しているデータ項目とそのメタデータの標準化を行う。

2. コホート横断検索システムカタログのアプリケーションの機能拡充とユーザビリティ向上

コホートと収集しているデータ項目のメタデータを検索するコホート横断検索システムカタログについて、ユーザから聞き取りを行い、アプリケーションの機能拡充とユーザビリティの向上を行う。

(倫理面への配慮)

本研究はヒトゲノム・遺伝子解析、臨床研究、ヒトを対象とする医学系研究、動物実験等の実施はない。したがって倫理面の問題はないと判断した。

C. 研究結果

1. コホート横断検索システムカタログの拡充のための大規模ゲノムコホートのデータ項目とメタデータの標準化

各コホート拠点より下記の通り、収集しているデータ項目とそのメタデータの提供を受け、標準化を行った。

- (1) 多目的コホート(JPHC)研究
JPHC 研究からは JPHC I、JPHC II、JPHC-NEXT のベースライン調査のデータ項目の提供を受け、標準化を行った。
JPHC I のベースライン調査(50,200 人)、JPHC II のベースライン調査(63,200 人)については、それぞれ 224 項目、279 項目の調査票のデータ項目であった。JPHC-NEXT のベースライン調査(115,385 人)については、調査票のデータ項目は 670 項目、検査のデータ項目は 49 項目であった。別途、調査票から計算された栄養データのデータ項目として 282 項目であった。
- (2) 日本多施設共同コホート研究(J-MICC)研究
J-MICC 研究からはベースライン調査(92,525 名)の 412 項目のデータ項目の提供を受け、標準化を行った。
- (3) 愛知県がんセンター病院疫学(HERPACC)

研究

HERPACC (Hospital-based Epidemiologic Research Program at Aichi Cancer Center)からは HERPACC2(28,773 人)、HERPACC3 (18,663 人)の調査の、それぞれ 304 項目、384 項目のデータ項目の提供を受け、標準化を行った。

(4) 神奈川県みらい未病コホート研究

神奈川県みらい未病コホート研究の調査(4,638 人/目標 5,000 人)からは 489 項目のデータ項目の提供を受け、標準化を行った。

(5) 慶應鶴岡メタボロームコホート(TMC)研究

慶應鶴岡メタボロームコホート(TMC)研究からはベースライン調査(11,002 名)の基本情報の2項目のデータ項目の提供を受け、標準化を行った。

これらのデータ項目について階層を整理し、標準化した。今後、コホート横断検索システムカタログへの収載を行う予定である。

2. コホート横断検索システムカタログのアプリケーションの機能拡充とユーザビリティ向上

コホートと収集しているデータ項目のメタデータを検索するコホート横断検索システムカタログについて、ユーザから聞き取りを行った。検索性能の向上、ユーザビリティ向上としてパーマリンク、ぱんくずリストの機能追加等があがった。検索性能の向上にあたってはバックエンドを改修し、ユーザビリティ向上としてはパーマリンク、ぱんくずリスト等のアプリケーションの機能拡充の改修を行った。

D. 考察

コホート横断検索システムのカタログを拡充するため、産業界からの要望が強いゲノムコホートから5拠点の大規模なゲノムコホートに対して、コホート横断検索システムカタログでの公開を前提として、各コホートで収集しているデータ項目とそのメタデ

ータの提供を依頼し、協力を得ることができた。また、ユーザから聞き取りを行い、アプリケーションの機能拡充とユーザビリティ向上を実現することができた。昨年度に収載したマイクロバイーム関連のコホートに続いて、ゲノムコホートを拡充できるようにしたことは、バイオ戦略の「世界最先端のバイオエコノミー社会」の三要素の一つ、バイオデータ駆動における「バイオ戦略のバイオとデジタルの融合」により、生物活動のデータ化等も含めてデータ基盤を構築し、それを最大限活用することにより産業・研究が発展するという目的に沿うもので、産業界がコホートのデータを利活用する基盤になると考えている。今回、協力を得ることができた5拠点のコホートのカタログについては公開を目指す、各コホート拠点が進めるデータ収集は拡大してゆくため、産業界がコホートのデータを利活用する基盤としてゆくためにも、本研究で構築した次世代バイオデータ基盤をいかに継続して運用し、発展させることができるかが重要である。

本研究により研究開発したコホート横断検索システムカタログを活用することで、コホート・バイオバンクの産業利用が促進することが期待される。コホート・バイオバンクの産業利用により、多様な分野において科学的エビデンスに基づいたサービスを提供できる環境を整備しデータ駆動型のヘルスケア産業を創出することにつながると考えられ、データを統合・強化する大規模健康人コホート・バイオバンクの構築を通じて、国民の QOL 向上に資する疫学研究の発展に寄与することが期待される。

E. 結論

次世代バイオデータ基盤の構築に向けて、産業界のニーズを踏まえて、コホート横断検索システムカタログへの5コホート拠点の拡充のためのデータ項目とメタデータの標準化と、アプリケーションの機能拡充とユーザビリティ向上を行った。産業界からの要望が強いコホートの拡充のため、調整・依頼を行い、のべ 373,384 人、2,811 項目ものデータ項

目・メタデータの提供を受け、標準化した。

F. 健康危険情報 なし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし