

厚生労働科学研究費補助金（政策科学総合研究事業）

（臨床研究等 ICT 基盤構築・人工知能実装研究事業）

総括研究報告書

リアルワールドヘルスデータのブリッジング標準化のためのシステム開発

～マイクロバイオーム分析データのブリッジング標準化～

国立研究開発法人医薬基盤・健康・栄養研究所

健康に対する意識が高まる中、腸内細菌に代表されるマイクロバイオームは健康に影響を与える重要な因子として注目されている。腸内細菌のゲノムデータの取得においては、便の回収から遺伝子の抽出、増幅、解読などゲノム情報を得るまでの手法がそれぞれ複数あり、また手法の違いが得られるデータに影響を与えることが分かっている。本事業では、健康との関連が注目されている腸内細菌を例に、異なる手法で取得したデータを同じデータソースとして有効活用できるようデータ変換によりブリッジングするシステムを開発し、標準化することを目的とする。

本年度は、本プロジェクト専任のデータブリッジングシステム開発チームを立ち上げ、知財・標準化検討委員会を設置した。本チームを中心に、データブリッジングの標準化に向けた検討会を開催し、アプリケーションの開発コンセプトを明確化し、システム開発に着手した。具体的には、医薬基盤・健康・栄養研究所で確立してきた NIBIOHN JMD (Japan Microbiome Database) などの研究基盤を活用し、異なるプロトコルで取得された遺伝子配列データについて、異なるデータベースや解析ソフトを用い、ブリッジングの計算に供するための菌叢データを取得した。各手法による解析結果への影響を把握し、実行可能性を検証するとともに、データブリッジングの標準化に向けた検討会など本システムのユーザーと想定される企業や大学などの研究者等と意見交換しながら、実際のブリッジングを行うためのアプリケーションの開発にも着手し、データブリッジングシステムの開発・標準化のための研究基盤を整備できた。

参画研究者：

米田悦啓（国立研究開発法人医薬基盤・健康・栄養研究所、理事長）

國澤 純（国立研究開発法人医薬基盤・健康・栄養研究所ワクチン・アジュバント研究センター、センター長）

水口賢司（国立研究開発法人医薬基盤・健康・栄養研究所 AI 健康・医薬研究センター、センター長）

宮地元彦（国立研究開発法人医薬基盤・健康・栄養研究所身体活動研究部、部長）

南里妃名子（国立研究開発法人医薬基盤・健康・栄養研究所身体活動研究部、室長）

寺内淳（一般社団法人日本マイクロバイオームコンソーシアム・運営委員会・委員長）

A. 研究目的

健康に対する意識が高まる中、個々の状態に応じた適切な介入により、健康を維持・増進できる社会の実現に向けた様々な取り組みが行われている。その一つとして、食生活などの生活習慣や健康状態に関する情報や、血液などの生体サンプルを用いて取得した生体機能に関するデータを用いたビッグデータ解析が進められ、「デジタルヘルス」や「デジタルトランスフォーメーション (DX)」というキーワードのもと、個人の生活習慣や健康状態のデータを活用するための基盤構築が進められている。

このような背景の中、健康に影響を与える因子として腸内細菌に代表される「マイクロバイオーム」が注目されている。ヒトの腸内には、約 1,000 種類、100 兆個の細菌が共生しており、健康との関わりが注目されている。古くから腸内細菌と健康との関わりは注目されていたが、従来の腸内細菌の解析は培養法を基本としており、9 割以上が培養困難な腸内細菌の解析を十分に行うのは困難であった。しかし、次世代シーケンサーを用いたゲノム解析の発展により、従来の培養法では詳細な解析が困難であった腸内細菌でも、ゲノムのデジタル情報から解析が可能となってきた。これらの発展に伴い、マイクロバイオームの知見を活用した創薬や機能性食品、ヘルスケアなど多くの産業で活用されることが期待されている。

腸内細菌のゲノムデータの取得においては、便の回収から遺伝子の抽出、増幅、解読などゲノム情報を得るまでの手法がそれぞれ複数あり、また手法の違いが得られるデータに影響を与えることが分かっていたことから解決方法として統一手法の開発が試みられていたが、個々の研究の目的に適

した手法があり、また既存データを継続的に使用するために手法の変更にも抵抗感があることから、手法の統一は世界的にも困難であるという認識になっている。

本事業では、健康との関連が注目されている腸内細菌を例に、異なる手法で取得したデータを同じデータリソースとして有効活用できるようデータ変換によりブリッジングするシステムを開発し、標準化することを目的とする。これにより、マイクロバイオームに着目したヘルスケア産業の発展を加速化させ、我が国のヘルスケア関連産業の市場拡大を図る。

B. 研究方法

倫理計画

ブリッジングデータを取得するための倫理整備・書類作成等を行った。具体的には、医薬基盤・健康・栄養研究所で確立してきた NIBIOHN JMD (Japan Microbiome Database)などのデータリソースの利活用、妥当性検証や高度化のための新規サンプルの収集、他コホートとの連携などデータブリッジングシステムの開発や最適化のための研究倫理ならびに体制を整備した。

16S rRNA 解析やショットガンメタゲノム解析による遺伝子配列データの取得

グアニジン塩を含む保存液を用いて採取した便や冷凍保管した便サンプルから、ビーズ破砕法や市販のキットなどにより DNA を抽出した。糞便 DNA から 16S rRNA 遺伝子の V3-V4 領域を PCR 法で増幅し、MiSeq を用いて塩基配列を解読し、Qiime パイプラインを用いて菌種を同定した。

同じ糞便 DNA を物理的に断片化、アダプターを付加し、NovaSeq6000 を用いて塩基配列を解読した。Bowtie2 を用いてヒトゲノ

ム由来の配列を除去し、マイクロバイオームのショットガンメタゲノム情報を取得した。

菌叢データの取得

16S rRNA アンプリコンシーケンシングデータは、Qiime パイプラインを基盤として作成したスクリプトを用いて SILVA や Greengene など複数のデータベースを参照し処理した。ショットガンメタゲノムシーケンシングデータは、Kraken2-Braken 法を用いた系統組成解析および FMAP 法を用いた遺伝子機能解析を実施した。

(倫理面への配慮)

本研究は、医薬基盤・健康・栄養研究所ならびに日本マイクロバイオームコンソーシアムにおいて倫理審査、承認を得た後、人を対象とする医学系研究に関する倫理指針に従って遂行した。

C. 研究結果

知財・標準化検討委員会の設置

知財や標準化に関する動向や企業ニーズを調査するため、JMBC 内に本プロジェクト専任のデータブリッジングシステム開発チームを立ち上げた。本チームは、製薬企業や検査企業、化学製品企業、研究受託企業など、マイクロバイオームに関連のある複数の業種の企業で構成されている。本年度は、医薬基盤・健康・栄養研究所と打合せを行い、企業の観点から標準化に向けた方向性を提案し、実現に向け各機関で担当する具体的な進め方や方法を決定した。

さらに、関連領域の他企業からマイクロバイオーム以外も含めて本コンセプトの可能性と期待についてヒアリングを行った。

倫理計画の作成と研究体制の構築

ブリッジングデータを取得するため、新しく「腸内環境の統合解析に関する研究」を立案し、倫理審査委員会にて承認を得た。具体的には、医薬基盤・健康・栄養研究所で確立してきた NIBIOHN JMD (Japan Microbiome Database)に格納されている遺伝子配列データなどの利活用が可能となり、データブリッジングシステムを開発し、さらに、本研究のために新しくヒトサンプルを収集し、データを拡充することで開発技術の妥当性の検証や高度化を行う体制を構築した。さらには、将来的に他の研究グループがすでに取得しているビッグデータを統合し解析するための倫理基盤として整備を進めた。

既存データの活用による情報解析手法の比較

日本人 28 名の 16S rRNA アンプリコンシーケンシングデータについて、世界的に汎用されている代表的なデータベースである SILVA と Greengene を用いて解析し、得られた菌叢データをもとに、異なるデータベースを用いた解析データについてブリッジング実行可能性を検証した。属レベルまで分類がされ、両データベースの結果に存在している細菌属を抽出し、Bray-Curtis 距離を用いて、サンプル間の腸内細菌叢の類似度を確認したところ、異なるデータベースであってもドミナント細菌の存在量は類似していた。この結果から、同じ手法で取得した遺伝子配列データであれば、データベースなど異なる情報処理により取得した菌叢データであっても互換性が高く、データブリッジングが可能であると考えられる。

このように、系統分類的に確立されている多くの細菌種についてはブリッジング可

能であるが、系統分類名が確立されていないなど一部の細菌種（割合も少ない）については、データベースごとで異なる名称で分類されているなど分類の定義が異なることが判明した。このような細菌種についてはブリッジングの際に注意が必要であり、今後、学名など基準を設定・変換するなどシステムを構築する必要があると考える。

既存データの活用による 16S rRNA アンプリコンシーケンシングとショットガンメタゲノムシーケンシングの比較

続いて、日本人 28 名の便サンプルについて、16S rRNA アンプリコンシーケンシングとショットガンメタゲノムシーケンシングで取得した菌叢データをもとに、異なる分析手法で取得したデータについてブリッジング実行可能性を検証した。16S rRNA アンプリコンシーケンシングのデータベースと比較と同様に、属レベルのデータについて Bray-Curtis 距離を用いて、サンプル間の腸内細菌叢の類似度を確認した。階層的クラスタリングや主座標分析の結果、16S rRNA アンプリコンシーケンシングとショットガンメタゲノムシーケンシングで取得した菌叢データは異なるクラスターを形成したことから、異なるシーケンシング手法で取得したデータは互換性が低いことが明らかになった。

異なる前処理手法（便の採取、DNA 抽出、ライブラリ調整など）でのデータ取得

JMBC ではこれまでに NEDO 事業にて糞便メタゲノム解析推奨プロトコルの開発を進めており、DNA や菌体カクテルを用いて 21 種の DNA 抽出方法、28 種のライブラリ作成方法を比較検討し、論文にてデータを公開している（Microbiome. 9:95 (2021),

Microbiol. Spectr. 10, e01915-21 (2022)）。本事業においても同じ DNA や菌体カクテルを標準品として使用、また、すでに取得済みの比較データを提供頂くことで了承を得ており、異なる DNA 抽出・ライブラリ調整手法で得られたデータのブリッジング検証に活用する。

また、糞便などの実サンプルから前処理の影響を検討するため、採便部位、タイミング、保管（保存液や冷凍など）、DNA 抽出（ビーズ破碎、酵素法、キットなど）など異なる条件でサンプルを収集し、データ取得を進めている。

また、マイクロバイーム解析の手法としては、次世代シーケンサーを用いたゲノム解析が主流であるが、フローサイトメトリーや質量分析などを用いた新しい手法の開発も進んでいることから、新規手法の動向調査やデータ取得も開始した。

アプリケーションの開発

上述した研究結果から、分析手法や解析方法によって互換性の高いケースと低いケースが存在することが想定される。また、データブリッジング手法についても、線形回帰などの単純モデルだけではなく、機械学習や AI などを活用したモデルなど、ブリッジングしたいデータによって、最適なアルゴリズムが異なると考えられる。また、データブリッジングの標準化に向けた検討会や企業や大学などの研究者等との意見交換なども踏まえ、本研究では、ユーザーがブリッジングするデータやアルゴリズムなどを自由に選択し、利用できるフレキシビリティの高いアプリケーションを開発することを念頭し、システム開発に着手した。

D. 考察

腸内細菌などのマイクロバイオームは、健康に影響を与える因子として注目されている。その種類や構成は食事などの生活習慣や遺伝的な要因の影響を受けると言われており、そのために個人差が大きいので、偶然誤差の考慮などのため大規模なデータによる研究が必要である。一方で、解析コストなどの理由から、1つの研究機関・グループでデータを取得することは現実的ではない。また、各研究機関が異なる手法を用いてデータを取得していることから、データ統合も難しく、比較的小さなコホートで取得したデータを用いた研究に留まっているのが現状の課題である。本研究において、ブリッジングシステムを開発し、標準化することで、異なる手法で取得したマイクロバイオームデータの相互利用が可能となり、データ連係による規模の拡大によって、データを最大限有効活用し、強固なエビデンスの創出や各種製品開発・産業創出の促進へつながると期待できる。ブリッジングシステムをより広く活用していくために、出口として産業界への展開も必要であり、本研究を共に行っている、JMBC と連携し推進していく予定である。

一方で、本年度の検討から、すべての手法で取得したデータを一律に統合できるのではなく、互換性の高いデータと低いデータが存在することが明らかになった。さらに、サンプル間の腸内細菌分布から菌の組成を予測することが機械学習的に実施し得る感触を得ている。これにより、データベース間で異なる名称がつけられていても、同一の菌ではないかといった検討が可能になる。また、ユーザーとなる研究者等との意見交換も踏まえると、研究者が自らの判断で、ブリッジングのためのデータやアル

ゴリズムなどを選択し、利用できるアプリケーションを開発することで、開発システムの利活用を促進し、マイクロバイオームの知見を活用した創薬や機能性食品、ヘルスケアなど多くの産業で活用される技術へと発展させたい。

また、医薬基盤・健康・栄養研究所では、これまでに健常人を主な対象としたデータ取得を行っており、現在までに日本国内の複数の地域の方を対象に 7,000 名を超える方のマイクロバイオームデータの収集を終え、独自に構築している NIBIOHN JMD データベースに格納しており、一部のデータを公開している。また、PRISM マイクロバイオームの支援を受けて「糖尿病個別化予防を加速するマイクロバイオーム解析 AI の開発」において、独自に開発している統合解析プラットフォーム MANTA を高度化し、個別化・層別化 AI システムの開発を進めている。これらの研究基盤とデータブリッジングシステムを連動させることで、健康・医療関連データをビッグデータ化ならびに健康に良い食の解明・開発とオーダーメイドな提供システムの開発などを通じて、市場領域を発展させるシステムを世界に先駆けて実現し、世界市場を獲得する土壌としていきたい。

E. 結論

医薬基盤・健康・栄養研究所の NIBIOHN JMD データベースなどを活用することで、データブリッジングのための予備検討を実施し、ブリッジングの実行可能性を確認した。解析結果に加えて、検討会やユーザーの意見などを踏まえ、アプリケーションの開発コンセプトを明確化し、システム開発に着手した。

F. 健康危機情報

なし

G. 研究発表

1. 論文発表

1. 河合総一郎、國澤純 マイクロバイオームから紐解く健康科学の近未来
日本歯科医学会記録集 (印刷中)
2. 細見晃司、國澤純 食と腸内細菌から考える健康長寿最前線 アンチ・エイジング医学 (印刷中)
3. 吉井健、細見晃司、國澤純 腸内細菌の代謝物を介した免疫機能制御 腸内細菌学雑誌 36(1): 1-11, 2022

2. 学会発表

なし

H. 知的財産権の出願・登録状況 (予定を含む)

1. 特許取得

該当事項なし

2. 実用新案登録

該当事項なし

3. その他

特記事項なし