

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

健康分野でのデータブリッジングシステムの国際標準化

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。

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中心とした知財・標準化検討委員会において、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。システム開発のためのデータ取得に際し、医薬基盤・健康・栄養研究所ならびに他機関において、本事業に関する倫理申請を行い、本事業に関連する各機関の倫理体制について情報を得た。各機関で承認を得た後、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、初版の作成を行った。

参画研究者：

中村祐輔（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）

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

### 倫理計画と研究体制の整備

昨年度までに、ブリッジングデータを取得するための倫理整備を行い、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が整備されており、今年度は、本倫理計画に他研究機関を追加し、研究体制を整備した。

### 16S rRNA 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。

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 データベースを参照し処理し、菌叢データを取得した。

### ブリッジングシステムの開発

ブリッジングシステムとして、菌叢データを俯瞰し、全体の群衆構造としてとらえ、それぞれの手法由来の菌叢をお互いに比較し、相手方の手法由来の菌叢を予測するアルゴリズムを開発した。これは、多数の説明変数から多数の目的変数を予測するため、多説明多目的予測のアルゴリズムである。モデルを評価するための評価指標として、真値と予測値の差異を示す RMSE 値を採用した。具体的なアルゴリズム例は以下の通りである。異なる菌叢の例として Shotgun 菌叢割合値と 16s 菌叢割合値を挙げるが、異なる手法で得られた同一人物の菌叢データであれば、手法に制限はない。

(1) Shotgun 菌叢割合値から 16s 菌叢割合値について機械学習的手法を用いて予測し、16s 菌叢割合実測値との誤差を求めた。

(2) 16s 菌叢割合値から Shotgun 菌叢割合値を機械学習的手法について予測し、Shotgun 菌叢割合実測値との誤差を求めた。

(3) 上記 (1)、(2) までで獲得したアルゴリズム計算式を用いることで、特定の菌について Shotgun から得られた菌存在割合値を入力すると、16s で取得した場合の菌存在割合値を予測し出力値として得ることができた。

(4) また逆に、特定の菌について 16S から得られた菌存在割合値を入力することで、Shotgun で取得した場合の菌存在割合値を予測し出力値として得ることができた。

(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 (JMBC) 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

#### 研究体制の構築

これまでに医薬基盤・健康・栄養研究所で確立してきた NIBIOHN JMD (Japan Microbiome Database)の利活用や新規ヒトサンプルの収集など、医薬基盤・健康・栄養研究所と日本マイクロバイオームコンソーシアム (JMBC) においてブリッジングデータを取得するための倫理計画「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ている。今年度は、データの拡充や開発技術の妥当性の検証や高度化のため、すでに腸内細菌研究を実施している多研究機関（早稲田大学、北海道大学、九州大学）においてブリッジングデータを取得するため、倫理計画を変更し、研究体制の構築を進めた。

現在の「人を対象とする生命科学・医学系研究に関する倫理指針」では、2022 年の改正に伴って代表機関で一括審査を実施することが可能となっているため、今回の倫理計画の変更について、医薬基盤・健康・栄養研究所の倫理審査委員会で一括審査を行い、他研究機関（早稲田大学、北海道大学、九州大学）で研究実施許可を得る予定であった。しかし、各大学の倫理審査委員会において本倫理指針が定める一括審査に対する体制の違いなどが判明したため、医薬基盤・健康・栄養研究所の倫理審査委員会で審査された後に、各大学の倫理審査委員会で改めて審査を受ける手順となり、当初の予定より時間と労力を要する結果となった。このような機関間における倫理審査

システムの違いは、ある程度は想像されることであったが、今後もデータ連携のための研究体制の構築において、研究機関間の研究倫理の審査体制や認識の違いが大きな課題となる可能性が明らかになった。

### ブリッジングデータの取得

これまでの既存データを用いた検討などから、分析手法や解析方法によってデータの互換性が高いケースと低いケースが存在することが想定されたことから、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開発を念頭にシステム開発が進められている。今年度は初版システムを用いたデータ連携の検証と最適化のためのブリッジングデータを取得した。具体的には、20名の被験者から異なる4つの手法にて糞便サンプルを複数

の機関において収集し、DNA調整、16S アンプリコンシーケンスによる遺伝子配列データを取得した（合計 80 サンプル）。具体的な手法と相違点は以下のとおりである。

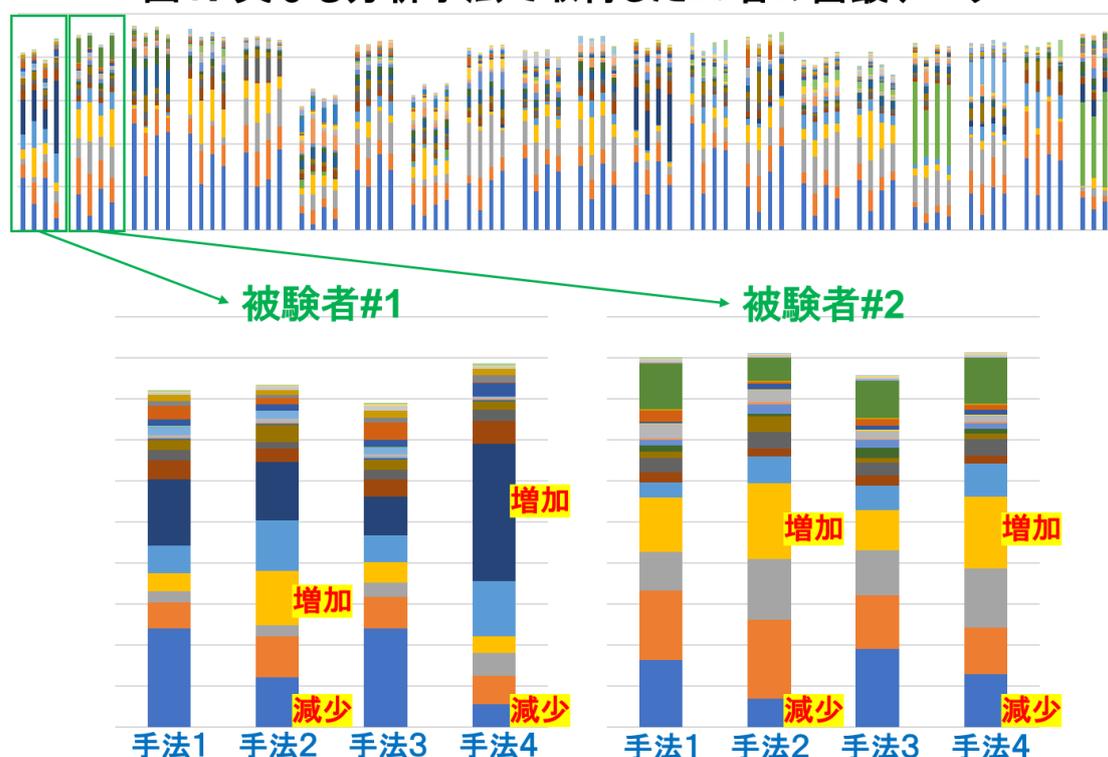
手法1：グアニジン溶液・スプーン型容器で採取、室温保管し、ビーズ破碎・核酸自動抽出器にてDNA調整した（NIBIOHNの手法）。

手法2：グアニジン溶液・スプーン型容器で採取、室温保管し、市販キットを用いてDNA調整した（DNA抽出手法が異なる）。

手法3：グアニジン溶液・ブラシ型容器で採取、室温保管し、ビーズ破碎・核酸自動抽出器にてDNA調整した（採取方法が異なる）。

手法4：グアニジン溶液を用いずに採取、冷凍保管し、ビーズ破碎・核酸自動抽出器にてDNA調整した（採取・保管方法が異なる）。

図1. 異なる分析手法で取得した20名の菌叢データ



## 異なる分析手法で取得した菌叢データの比較

今回の 20 名の被験者から取得した 80 サンプルの菌叢データにおいて、208 属の細菌が同定された。まず、分析手法の違いによって検出される菌と検出されない菌が存在するのか検討したところ、例えば、手法 1 と手法 2 では 170 属がどちらの手法でも共通して検出され、10 属が手法 1 でのみ、14 属が手法 2 でのみ検出された。しかし、どちらかの手法で検出された 24 属はいずれも相対存在量の平均値が 0.01%未滿の細菌であり、16S アンプリコンシーケンス法において検出限界の下限域の非常にマイナーな細菌属であった。さらに、その他の手法でも比較したところ、208 属の中で相対存在量の平均値が 0.01%を超える 117 属は、すべての手法において検出されていたことから、少なくとも今回検討した 4 つの手法においては分析手法によって検出できない細菌はデータブリッジングに影響を与えない誤差範囲であると考えられた。

分析手法で取得したデータにおけるサンプル間の腸内細菌叢の類似度を確認したところ、手法 1 と手法 3 で取得した菌叢データは非常に類似しており (図 1)、割合の多い 30 属の細菌について相関係数を算出したところ、23 属が 0.9 以上と非常に高い相関が示された。一方で、手法 1 と手法 2 や手法 1 と手法 4 では菌叢データに差異が認められ (図 1)、また、0.9 以上の相関となったのはそれぞれ 21 属と 14 属であり、手法 1 に対して手法 3 は互換性が高く、手法 2 は中程度、手法 4 は互換性が低いことが示唆される結果であった。現在、これらのデータをもとに初版システムにおけるデータブリッジングの高度化を進めている。

## アプリケーションの開発と検証

これまでに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、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る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日本マイクロバイオームコンソーシアム (JMBC) 内に立ち上げ、知財・標準化検討委員会を設置している。想定利用者である産業界等の意向などに関する調査結果をもとに、開発コンセプトを「すべてのコホートがデータ連携したデータベースではなく、個々のコホートのユーザーが任意に連携先を選択できるシステムとする」と明確化し、ユーザーフレキシビリティの高いアプリケーション開発を進めている。

今年度は特に

- ・アプリケーションのシステム概要
- ・開発の進捗状況

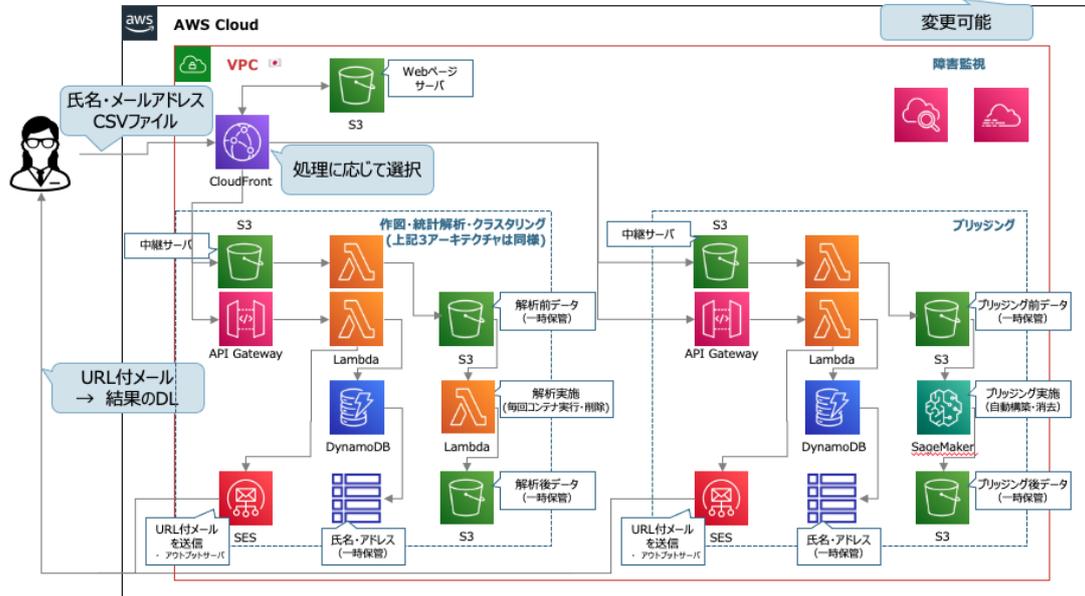
複数拠点のデータ受け入れる環境として、Web ページで作図、統計解析、クラスタリング、ブリッジング等の機能を選択してボタンを押すと各処理が実行される「(1) 初心者向け環境 (図 2)」と、Jupyter notebook で高い自由度でデータの前処理とブラッシュアップを行える「(2) アドバンスド環境 (図 4)」の二つを構築した。

### (1) 初心者向け環境 (図 2)

本環境では、利用者はまずサイトマップ (図 3) の PC-01 で氏名と結果返却用の E-mail アドレスを登録する。登録したメールアドレスには解析ページ (PC-02) へアクセスできる URL を記載したメールが送信される。登録用 API の連打や DDoS 攻撃対策として、API Gateway へのリクエスト回数が閾値を超えると、リクエストを拒否し、ユーザーにエラーメッセージを送る機能が実

## 図2：初心者向け環境詳細アーキテクチャ

■ 解析結果等返却用のために氏名（宛名用）、メールアドレスを登録してもらふ。（メール送信後、30分で消去）



装されている。解析ページ(PC-02)では、利用者が各種の解析機能を選択でき、「作図(PC-03)」、「統計解析(PC-04)」、「クラスタリング(PC-05)」で、自身で用意したデータ、あるいはサイト上に用意される NIBIOHN のデモデータを参照してから、実行ボタンを押すことで、利用者が解析結果の出力される S3 バケット（クラウドストレージ）に一定時間アクセスでき、結果を取得することができる。

「データブリッジング（PC-06）」では、利用者は異なる方法で取られた 2 つのデータ（自身で用意したデータあるいは、NIBIOHN 公開データから選択可能）を選択したのち、ブリッジングメニューから実行する推論方法を選択して実行ボタンを押す。利用者にはブリッジング結果に一定時間アクセスできる期限付き URL の記載されたメールが自動送信される。アップロードされたデータはデータ入力用の AWS S3 バケットに送信され、それを発火点として、学習済みの機械学習モデルをホストした推論実行サー

バーが起動し、ブリッジングが実行される。結果は利用者が期限付き URL でダウンロードできるようになる。

ブリッジングの結果を用いて、作図、統計解析、クラスタリングを行うことで、推論の精度確認や不要なデータのトリミング等を行い、再度ブリッジングを実行することも可能となる。

### (1) アドバンスド環境（図 4）

本環境は大きく分けて「前処理・ブラッシュアップ環境」と「ブリッジング環境」の 2 つの環境からなり、どちらを利用するかは Web サイトで選択できる。

「前処理・ブラッシュアップ環境」では、利用者が氏名と結果返却用の E-mail アドレスを登録すると、前処理実行環境へ一定時間アクセスできる期限付き URL がメールで自動返却される。前処理実行環境では、利用者は Jupyter notebook 上でコードを実行することにより、データの要約統計量の表示やグラフによる可視化、クラスタリン

図3：サイトマップ（初心者向け環境ワイヤフレーム）

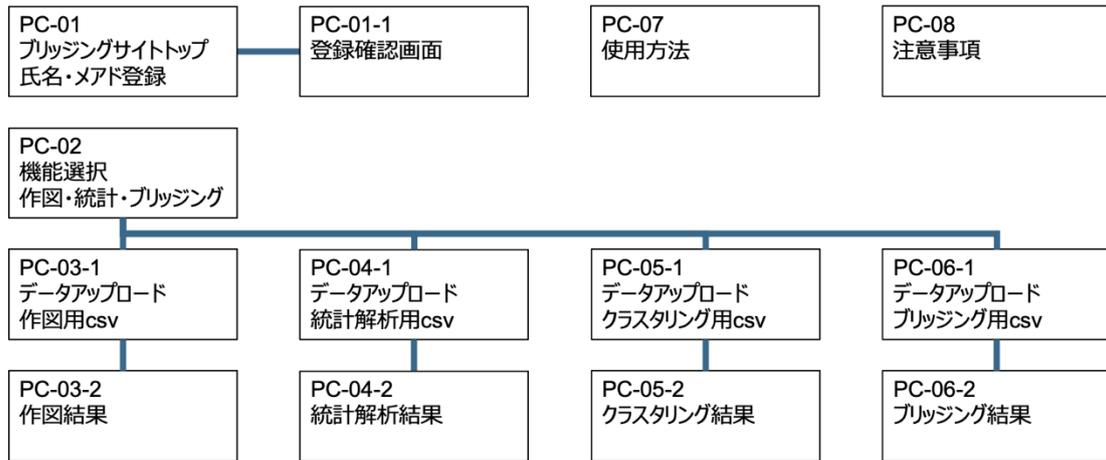
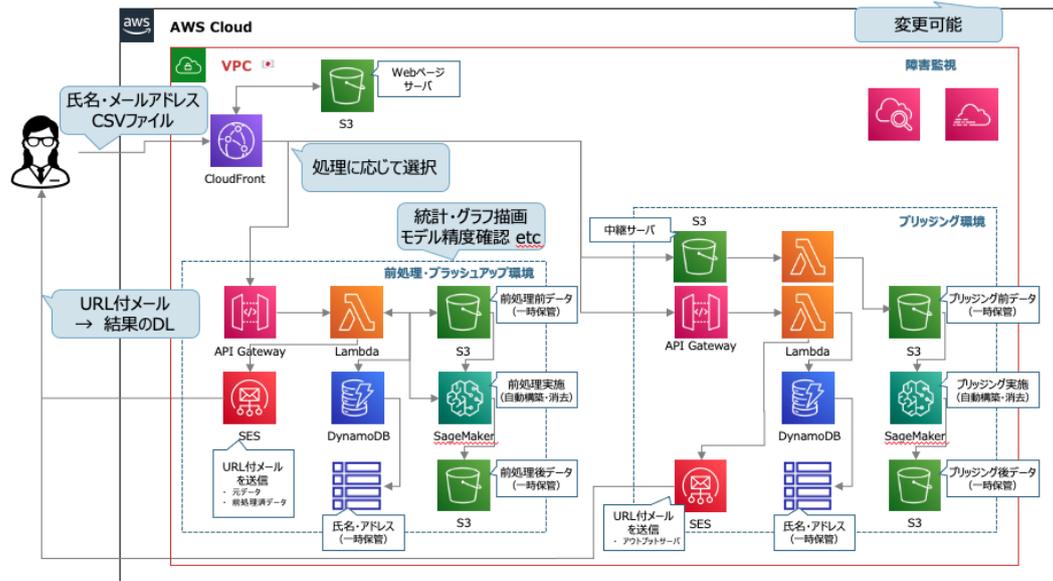


図4：アドバンスド環境詳細アーキテクチャ

■ 解析結果等返却用のために氏名（宛名用）、メールアドレスを登録してもらう。（メール送信後、30分で消去）



グなどが可能である。自動返却メールには、前処理用データをアップロードしてクラウド内で一時的に保管・管理できる S3 バケットと、前処理後のデータを同じくクラウド内で一時的に保管・管理できる S3 バケットにアクセスするための URL も記載される。また、この「前処理・ブリッジング環境」は、後述するブリッジング環境で得られた推論結果を可視化・モデル精度確認、クラスタリングする目的でも利用可能されることを見込んでいる。「ブリッジング環境」に

おいても、利用者が Web サイトに氏名と結果返却用の E-mail アドレスを登録するところまでは同様である。「ブリッジング後データ」が入る S3 バケットに一定時間アクセスできる期限付き URL が自動返却される。同時に、Web サイトから機械学習サーバーへデータを中継する S3 バケットにデータをアップロードできるようになる。S3 バケットにアップロードされたデータは、自動的に学習済みブリッジングモデルがホストされた機械学習サーバーへと送信されそこで

推論が実施される。推論結果は、CSV 形式でブリッジング後データ S3 バケットに送信され、利用者は前述の期限付き URL で推論結果ファイルをダウンロードできる。

・異なる分析手法で取得した菌叢データを用いた検証結果

現在、異なる分析手法でのデータの受領が進捗している。実際の菌叢データについて、いずれかの被験者において 0.1%以上の存在比がある菌を全て用いてお互いの菌数相互予測モデルを XGBoost (default status) にて作成したとこと、良好な検証結果を得ている (RMSE 値 $\leq$ 1.0)。一方でゼロ過剰データである菌叢データの相互予測においては、0 から 0 を予測するケースが多く、これが精度の低下につながっている。一方で、実数値を持つ菌叢データについての検証には改善の余地があり、ドメイン知識を用いた適切な菌の選択といった工夫が必要である。

#### 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おり、遺伝学的研究手法が主流となっているが、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、解読など分析手法が複数あり、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知られている。そこで、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、データ連携や最大限の有効活用、強固なエビデンスの創出、各種製品開発・産業創出の促進へつながると期待される。関連事業として、昨年度は、一般社団法人日本マイクロバイ

オームコンソーシアム (JMBC) 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を開催し、本事業の開発コンセプトを「すべてのコホートがデータ連携したデータベースではなく、個々のコホートのユーザーが任意に連携先を選択できるシステムとする」と明確化しており、本年度は本コンセプトに基づいたシステム開発と標準化を進めた。

実際にデータ連携を進める上での課題の洗い出しのため、腸内細菌コホートをされている他研究機関 (早稲田大学、北海道大学、九州大学) の協力を得て、データ連携を進める上での課題の洗い出しを進めた。その中で、研究機関間の研究倫理の審査体制や認識の違いが大きな課題となることが明らかになった。これは各機関での考え方の違いもあり、各機関を一律の形で進めることが決めることができなくなる大きな要因の一つとして十分に考慮すべき点であると言える。

アプリケーション開発においては、異なる手法を用いて取得した菌叢データを用いて初版システムの検証と高度化を進めており、今後、実際に多施設で取得したデータや既存コホートにおけるデータ連携を検討し、倫理計画なども含めたパッケージとしてアプリケーションの実用化を進め、マイクロバイオーム領域をはじめとする健康・医療関連データの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く。

#### E. 結論

「すべてのコホートがデータ連携したデ

ータベースではなく、個々のコホートのユーザーが任意に連携先を選択できるシステムとする」の開発コンセプトのもとに、フイージビリティスタディを行い、倫理計画などデータ連携を進める上での課題を洗い出すとともに、菌叢データをブリッジングできるシステムを開発した。今後の本システムパッケージの実用化を進め、世界のデータ駆動型研究をリードする技術基盤を構築したいと考えている。

#### F. 健康危機情報

なし

#### G. 研究発表

##### 1. 論文発表

1. 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)
2. 國澤純 「腸内細菌研究から見てきた健康科学の最前線」によせて **FFI ジャーナル** 227(3): 197-200, 2022
3. 細見晃司、國澤純 ポストバイオティクス研究から考える健康科学の将来と個別化栄養 **Food Style**21 6: 64-67, 202
4. 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 歯科展望特別号 逆転の発想 歯科界 2040年への挑戦 143-144, 2022
5. 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
6. Sawane K, Hosomi K, Park J, Ookoshi K, Nanri H, Nakagata T, Chen YA, Mohsen A, Kawashima H, Mizuguchi K, Miyachi M,

Kunisawa J (2022) Identification of human gut microbiome associated with enterolignan production.

*Microorganisms*10(11): 2169. doi: 10.3390/microorganisms10112169.

7. Mohsen A, Chen YA, Osorio RSA, Higuchi C, Mizuguchi K (2022) Snaq: a dynamic snakemake pipeline for microbiome data analysis with QIIME2. *Front Bioinform* 2: 893933. doi: 10.3389/fbinf.2022.893933.
8. Hosomi K, Saito M, Park J, Murakami H, Shibata N, Ando M, Nagatake T, Konishi K, Ohno H, Tanisawa K, Mohsen A, Chen YA, Kawashima H, Natsume-Kitatani Y, Oka Y, Shimizu H, Furuta M, Tojima Y, Sawane K, Saika A, Yonejima Y, Takeyama H, Matsutani A, Mizuguchi K, Miyachi M, Kunisawa J (2022) Oral administration of *Blautia wexlerae* ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat Commun* 13(1): 4477. doi: 10.1038/s41467-022-32015-7.
9. Park J, Hosomi K, Kawashima H, Chen YA, Mohsen A, Ohno H, Konishi K, Tanisawa K, Kifushi M, Kogawa M, Takeyama H, Murakami H, Kubota T, Miyachi M, Kunisawa J, Mizuguchi K (2022) Dietary vitamin B1 intake influences gut microbial community and the consequent production of short-chain fatty acids. *Nutrients* 14(10): 2078. doi:

- 10.3390/nu14102078.
10. Otoshi T, Nagano T, Park J, Hosomi K, Yamashita T, Tachihara M, Tabata T, Sekiya R, Tanaka Y, Kobayashi K, Mizuguchi K, Itoh T, Maniwa Y, Kunisawa J, Nishimura Y (2022) The gut microbiome as a biomarker of cancer progression among female never-smokers with lung adenocarcinoma. *Anticancer Res*42(3): 1589-1598. doi: 10.21873/anticanres.15633.
  11. Ikubo Y, Sanada T, Hosomi K, Park J, Naito A, Shoji H, Misawa T, Suda R, Sekine A, Sugiura T, Shigeta A, Nanri H, Sakao S, Tanabe N, Mizuguchi K, Kunisawa J, Suzuki T, Tatsumi K (2022) Altered gut microbiota and its association with inflammation in patients with chronic thromboembolic pulmonary hypertension: a single-center observational study in Japan. *BMC Pulmonary Medicine* 22(1): 138. doi: 10.1186/s12890-022-01932-0.
  12. Yoshimura E, Hamada Y, Hatanaka M, Nanri H, Nakagata T, Matsumoto N, Shimoda S, Tanaka S, Miyachi M, Hatamoto Y, (2022) Relationship between intra-individual variability in nutrition-related lifestyle behaviors and blood glucose outcomes under free-living conditions in adults without type 2 diabetes. *Diabetes Research and Clinical Practice*
2. 学会発表
    1. 國澤純、腸内環境が作り出す代謝パスイェイを活用した健康科学への新展開 **レドックス R&D 戦略委員会** (九州大学・福岡、2023年3月17日) 招待講演
    2. 國澤純、腸内環境研究から創出する新しい創薬・食品モダリティへの展開 **日本農芸化学会 2023 年度大会スポンサーセミナー** (オンライン、2023年3月15日) 招待講演
    3. 國澤純、腸内環境研究から創出する新しいヘルス・メディカルサイエンスへの挑戦 **共創の場オールスター最先端セミナー第 10 回** (オンライン、2023年1月24日) 招待講演
    4. 國澤純、近未来型健康社会の実現に向けた腸内環境の理解と応用 **抗加齢医学会 Web メディアセミナー** (オンライン、2022年11月15日) 招待講演
  - 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  
該当事項なし