

協調可能なバイオデータ等を活用した試行的取組

研究分担者 熊田 和貴 東北大学東北メディカル・メガバンク機構バイオバンク部門 教授

研究要旨

産業界とコホート拠点との連携促進を図ることを目的に協調可能なバイオデータ等を活用した試行的な取組をおこなった。コホート・バイオバンクの潜在的な需要が見込まれる 40 社を超える企業にアプローチし、本 PRISM 事業で構築したコホート横断検索システムや窓口機能について周知するとともにコホート・バイオバンクの活用方法についての説明をおこない、その積極的な利活用を促した。その結果、約 4 割の企業がコホート・バイオバンクを活用したハッカソンや共同研究などを検討する段階へと進んだ。それらの企業に対して研究開始に必要な研究計画立案や倫理申請手続き等のコーディネーションをおこない、窓口機能の有効性を確認した。また、一部企業に対しては 2022 年度中に産業界とのデータ利活用勉強会やハッカソンを行うことを計画し、その実施に向けて解析環境の整備等をおこなった。本分担研究により産業界のコホート・バイオバンクに対する高いニーズが改めて確認され、これらの企業が実際にコホート・バイオバンクの利活用を経験することで、今後より積極的かつ効率的な産学連携研究開発の実施につながると期待される。

A. 研究目的

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に向けて、産業界によるコホート・バイオバンクの利活用を介した科学的なエビデンスに基づくデータ駆動型のヘルスケア産業の創出が望まれている。産業界からもコホート・バイオバンクが保有するデータ等に対するニーズは高いが、その一方で実際の利用にあたっては様々なハードルが存在し積極的な利活用が進みにくい状態であることが指摘されている。本 PRISM 事業においても 2019 年度から 2020 年度までの調査研究によって、個々のコホート・バイオバンクに格納されているデータ内容等を一括で検索できる横断検索システムをはじめとしたシステム面での利便性確保、関連法規等制度面での対応の必要性、個別の開発ニーズに応じた支援の必要性といった課題を抽

出し、これらの解決を目指して 2020 年度から 2021 年度にかけて試行的な横断検索システム及び窓口機能の構築を行ってきた。本分担研究ではこれらのコホート・バイオバンクの産業利用促進策の具体化を踏まえ、協調可能なバイオデータ等を活用した試行的取組を介して、それらの機能を実証しつつ産業界とコホート拠点との連携促進を図ることを目的としている。

B. 研究方法

産業界とコホート拠点との連携促進を図るために協調可能なバイオデータ等を活用した試行的取組として、1.コホート拠点と産業界とのマッチング、2.産業界とコホート拠点との共同研究に係る研究計画立案のコーディネーション、3.解析環境の構築を含む共同研究の試行に取り組んだ。

コホート拠点と産業界とのマッチングにおいては産業界を代表する組織として日本製薬工業協会ならびに関西医薬品協会との意見交換をおこなった。またコホート・バイオバンクの潜在的な利活用が見込まれる40社を超える個別企業に対しても働きかけをおこない、本PRISM事業で構築したコホート横断検索システムや窓口機能について周知するとともにコホート・バイオバンクの活用方法についての説明をおこない、その積極的な利活用を促した。

産業界とコホート拠点との共同研究に係る研究計画立案のコーディネーションにおいては、コホート・バイオバンクの利活用を計画している企業に対して、本PRISM事業で構築した窓口機能の実証も兼ねて、研究開始に必要な研究計画立案や倫理申請手続き等のコーディネーションをおこなった。

解析環境の構築を含む共同研究の試行においては、実際に産業界が利活用できるデータとして東北メディカル・メガバンク機構(ToMMo)が2021年度SIP事業に参画して収集した300人分の糞便検体のマイクロバイオームデータを準備し、ToMMoが有する既存の情報と合わせて産業界との試行的解析(データ利活用勉強会)や共同解析(ハッカソン)を計画している。2022年度中の実施に向けて解析環境の整備等をおこなった。

なお、本分担研究は部分的に株式会社ちとせ研究所に委託された「次世代バイオデータ等の試行的産業利用に関する支援業務」の一環として実施されている。

(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. コホート拠点と産業界とのマッチング

ヘルスケアデータを有するコホート拠点との共同研究等取り組みのマッチングを意図した各種活動

を実施した。具体的には①日本製薬工業協会との意見交換②関西医薬品協会との意見交換③個別企業へのアプローチに大別される。

(1) 日本製薬工業協会との意見交換

2021年2月19日に一般財団法人バイオインダストリー協会(JBA)で開催された第36回ヘルスケア研究会 総会・講演会にてちとせ研究所が講師を務めたことをきっかけとして、バイオ戦略を実現するためのPRISM事業の成果を踏まえた産業界によるヘルスケアデータ利活用を促進するための意見交換を実施した。意見交換は製薬協 研究開発委員会 産学官連携部会の開催と併せて実施され、産学官連携部会長、研究振興部長、専門委員長が現地参加し、産学官連携部会以外の研究開発委員会メンバーを含む29名がWebから参加した。この意見交換の結果、データ解析、知的財産、中間管理事業それぞれについて、活発な議論がなされた。拠点間連携をすすめ、それらデータや仕組みを認定し、データ解析が可能になるような枠組みが必要であると合意した。今後の施策として、データ利活用を加速するために特に必要な機能は拠点ごとのデータマネジメントプラン・前処理済みデータの整備と公開であると結論づけた。以下に、今後実施すべき、当日の議論で挙げられた論点を示す。

<データ解析>

・解析サービスについて、データサイエンティストの質や、解析結果の妥当性を担保する仕組みが必要だと考える。個別のプロジェクト独自の基準ではなく、データを連携し、使いやすいデータや解析基盤だけでなく、そこから得られた結果・解釈の妥当性・精度が判断できるような仕組みが必要。

・データ前処理がとて大変であり、この前処理がしっかりできていることが利活用の鍵を握る。同義語や、オントロジーの整理がなされているとよい。海外のある拠点で、前処理がしっかりなされているデータセットの存在がある。

・データサイエンティストが足りないとはよく言われているが、具体的にそのステージで足りないのか、ブレイクダウンして把握するとよいと考えられる。

<知的財産>

・NDB(疾患レセプト)について、データ利用申請手続きに半年、解析に1年かかったケースがある。二次利用のためには、適切なサポートがないと利活用が進まない。

・知財の考え方については、学校・研究法人によって違う。例えば、東北大のように分譲することを前提としているところがある。一次利用以外で利用することを前提として考えないと活用が進まない。

・大学側が制限し、囲い込み、先生方が営利企業に使わせたくないという壁を感じる。ToMMoのように計画時から二次利用を見据えた設計が必要。

・自分たちの論文を書くために3～5年は外に出てこないというケースがあり、これではデータが陳腐化してしまう懸念がある。

・データマネジメントプランをデータ取得時に定め、自分たちが研究に使いたいスコープ(一次利用目的)はデータ取得者が優先的に利用する権利を認めた上で、それ以外の部分(二次利用目的)についてはできるだけ早い段階で交渉に応じるべきという取りきめが必要。

・データ取得者は、自身の関心事である一次利用目的以外の利用を推進する事に対するインセンティブが働かないので、二次利用を推進するためには強力な後押しが必要。

<中間管理事業>

・大学の産学連携は、大学の先生方の成果と社会を橋渡しする産学連携室を持っている。一方で産業界の足りないところのサポートをすることを目的とした中間管理事業者が必要である。

(2) 関西医薬品協会との意見交換

ToMMoと株式会社ちとせ研究所は2022年2月18日に第1回研究開発推進会議(部会)での講演と意見交換を実施した¹。

議論の内容は以下の3点であり、19社29名が参加した。参加者からの事後アンケートの結果を表1に示す。

○バイオ戦略で謳われているヘルスケアデータ利活用に向けた論点について、内閣府のSIP²、PRISM事業の概要・取り組み(2019～2021年度)と併せて論点を整理して産業界の意向を汲み取ることがを目的とした議論

○ToMMoでの取り組み(横断検索システム・企業向け相談窓口)およびToMMo視点でのPRISM連携・将来像を紹介してコホート拠点の描く将来像を啓蒙

○実際の取り組み(ハッカソン(共同解析)、事業開発PoCプロジェクトの提案)

表1. 関西医薬品協会との意見交換後に寄せられた意見

ヘルスケアデータの利活用推進に向けた議論
・(ToMMoでは、)15万人もリクルートできたことは素晴らしいことだと思います。大阪では、何かインセンティブがなければ特に健康人のデータを入手することは難しいと思います。
・得られた貴重なデータは、是非、有効活用していただきたいと思います。
・オープンに出きる利活用例が知りたかった。
・課題抽出が行われ、利活用の絵姿も描かれており、議論は生産的で期待できるようにも感じられるが、実際は悲観的で、今までのデータで使えるのは東北MMBのみ。これからのビッグデータに期待。マイクロバイオームのデータの活用を

¹ <http://www.kpia.jp/osirase/202201261.html>

² 戦略的イノベーション創造プログラム(SIP)第2期 スマートバイオ産業・農業基盤技術

成功例として示せないか希望します。

東北メディカル・メガバンク機構での取り組みと
コホート拠点の描く将来像

・素晴らしい活動と思います。個人情報の問題も
クリアし、産業界が利用可能な優れた DB になる
ことを期待するとともに、国内外で拡張して欲し
い。

・データ収集のフェーズから利活用のフェーズ
に移行されていることが良くわかりました。また、
ToMMo が日本の一大拠点となっていることも理
解できました。

・東北メディカル・メガバンクについて、その経緯
とめざすところを大変わかりやすくご説明頂き、
有難うございました。これだけの産業利用も可能
な貴重なデータが無駄にしないよう、利活用を
進めていければと思います。

・取組みには敬意を表しますし、東北 MMB のデ
ータ活用には大いに期待します。東北の参加者
とは違って、他のコホートはなかなか難しいよう
に感じ、住民が楽しんで参加できるような仕組み
が必要と思った次第です。

ハッカソン(共同解析)、事業開発 PoC プロジェ
クト

・苦勞されながらも、産官学の連携を取りつつ、
実働させようと動きつつあることが良くわかりまし
た。

・ちとせ研究所のめざすところ、東北メディカル・
メガバンクでのデータ解析概要につき、大変わ
かりやすくご説明頂き、有難うございました。この
ようなデータ解析を専門とするとところとのコラボの
重要性を痛感しました。

・今後は、特に未病の段階から病気にならない
ように健康管理していくことが重要となってくる
。地域ごと、個人別の健康管理に結び付けて
いければと思います。

・ちとせ研究所の提供がなければなかなか難し
い共同解析をよく実施したものだと思います。

【3】御社が興味のあるヘルスケア領域について（※複数可）

回答パターン



【4】御社がヘルスケアビジネスを展開するうえで興味のある分野について（※複数可）

回答パターン



意見交換から得られた、今後実施すべき事項をまとめると以下の4点に集約される。

- マイクロバイオームのデータの利活用を題材として、産業界・社会にオープンにできる成功事例を提示する必要がある
- 産業界が利活用可能な横断検索システムの構築を拡張し、国内外のデータベースを網羅できる必要がある
- 蓄積されたデータが有用であり、利活用に対して期待感が高まっている一方で、全国レベルで見た時に、住民が前向きに参加できるような仕掛けが必要である
- いわゆる未病の段階からの健康管理につながるようなサービスの構築が求められている

(3) 個別企業へのアプローチ

コホート・バイオバンクの潜在的な需要が見込まれる企業を選定し、延べ40社の個別の企業に対してコホート・バイオバンクについての詳しい説明をおこなった。またその際、本PRISM事業で構築したコホート横断検索システムや窓口機能に関しても周知するとともに、それらを利用したコホート・バイオバンクの積極的な利活用を促した。また、これとは別に、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)にて2021年11月17日に開催された第5回企業間交流会において、約70名の参加者に対して周知を実施した。

これらアプローチの結果、元施策を実施している医薬基盤・健康・栄養研究所(NIBIOHN)との共同研究1件が開始され、ToMMoの分譲データを用いた共同研究1件が2022年度に開始される予定である。これら共同研究は従来型の産学官連携にとどまらず、外部協力機関(医療機関・自治体等)を巻き込んでいることが特徴であり、単にデータ解析では終わらない、データ駆動型の新規事業開発を目的とした民間投資による取組である。付随して、

株式会社ちとせ研究所と、国立遺伝学研究所との間で、国立遺伝学研究所が開発を進めているMicrobeDB.jpとの社会実装を目的とした連携が構築された。この連携によって、SIP事業においてToMMoによって取得された糞便のマイクロバイオームデータの利活用が加速される素地が整った。

さらに、後述するSIPハッカソンが8社の民間投資で進捗中であり、ビッグデータ解析における適切な次元削減・層別化が課題となっている。このSIPハッカソンでは、国立研究開発法人科学技術振興機構(JST)バイオサイエンスデータベースセンター(NBDC)のNBDCグループ共有データベースから公開³されている調査票、各種バイタルデータ、睡眠・心拍等デバイスデータと腸内細菌叢データを用いた解析が実施されている。参加企業のデータ解析スキルの向上が著しく、適切な成果発信と引き続きの企業へのアプローチを踏まえて2022年度に20社程度の民間投資が期待できる(図1、図2、図3)。

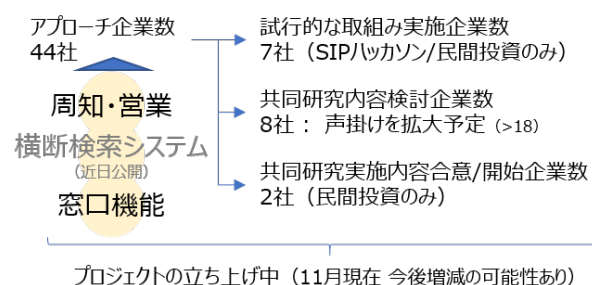


図1. 試行的産業利用全体図

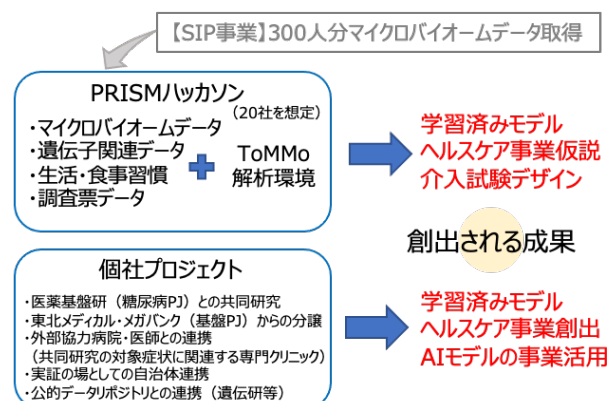


図2. 個別企業へのアプローチ

³ <https://gr-sharingdbs.biosciencedbc.jp/shd>

民間投資の拡大

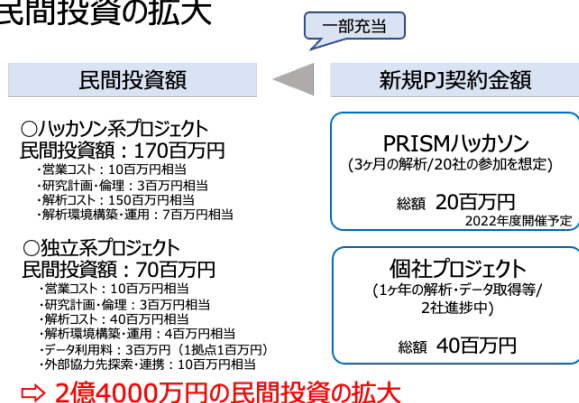


図3. 2022年度の民間投資拡大見通し

2. 産業界とコホート拠点との共同研究に係る研究計画立案のコーディネーション

上述の個別企業へのアプローチによってコホート・バイオバンクの利活用を検討するに至った企業のほか、ToMMoの有する試料・情報の利活用を計画して独自にToMMoにコンタクトをとった複数企業を対象に、本PRISM事業で構築した窓口機能を活用して、研究開始に必要な研究計画立案や倫理申請手続き等のコーディネーションをおこなった。窓口機能の充実により、コホート・バイオバンク側の労力の大幅な低減や、各手続きに要する期間の短縮が実証された。企業側の満足度の向上や計画から研究開始までに要する期間全体の大幅な短縮が見込まれる。

また、窓口業務のうちコホート・バイオバンク側が最低限有さなければいけない機能と条件によっては民間のサービサーが担いする役割についての検討もおこなった。産業界がコホート拠点と共同研究を開始する場合、リサーチクエッションの設定や研究計画の策定からデータ利用に至るまで様々なハードルが存在する。本PRISM事業において、データ利活用のための中間管理事業の必要を提言し、実証を進めてきた(図4)。そこで、SIP事業(2B)の社会実装として、一般社団法人の設立が進んでい

るが⁴、役割分担の検討のための意見交換を実施した(図5、図6)。結果として、役割分担については線引きができ、互いの役割を補完する形でデータ利活用を促進できる体制が構築され得ることを確認した。

窓口機能の機能と分担(試行)

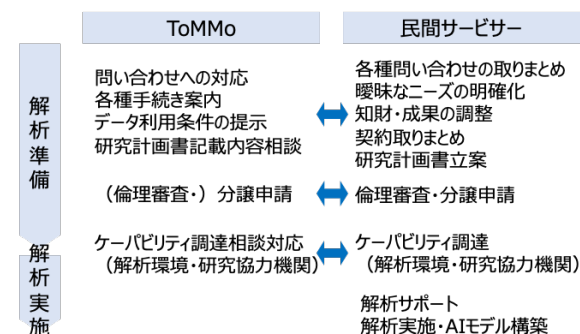


図4. 東北メディカル・メガバンクにおける窓口業務の役割と、補完する民間サービサーの役割

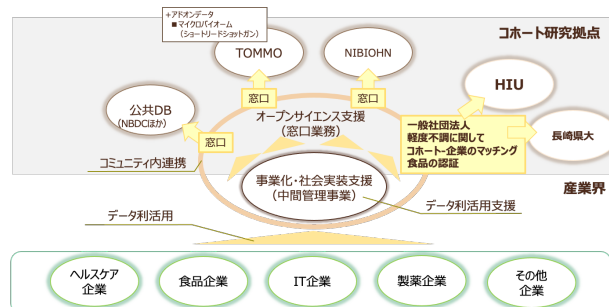


図5. 中間管理事業の位置づけ

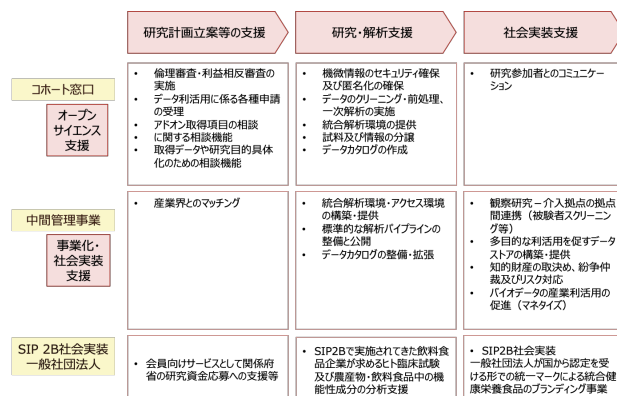


図6. データ利活用から社会実装に向けた役割分担の例

⁴ 一般社団法人日本セルフケア推進協議会

3. 共同研究の試行(解析環境の構築を含む)等

これまで検討・実証を進めてきた役割分担の実証の場として、SIP 事業における民間投資の拡大を同時に実現する SIP ハッカソンの企画・運営を通じたデータ利活用の試行的利用を実施した(図 7)。ToMMo が 2021 年度 SIP 事業に参画して収集した 300 人分の糞便検体のマイクロバイオームデータを実際に企業が直接利活用できるデータとして用意している。解析対象データは NBDC が認める機関外サーバーとしての国立遺伝学研究所が提供するスーパーコンピュータ上のストレージにて管理し、データ保管・転送・解析・結果出力プラットフォームを同スーパーコンピュータ上に構築した。これらのセキュリティは国立遺伝学研究所が提供するスーパーコンピュータ上で十分に管理されている(図 8)。この解析環境構築を好例として、SIP-PRISM 連携としての解析環境の ToMMo スーパーコンピュータシステム上への構築が実現した(図 9)。

また、データ解析においては、多大な労力が必要とされるデータ前処理から機械学習モデル作成を協調領域として設定した。ここでは、全くのデータ解析初心者が応用的な統計的因果推論、因果探索、ベイジアンネットワークの構築といった従来の古典疫学的な解析(食品 XX の摂取は疾病リスク YY の低下と関連が認められる 等)に止まらず、ドメイン知識の投入による因果関係の仮説構築や、生活習慣改善リコメンドに繋がる機械学習モデルの構築が実現している。

これら実例を踏まえて、ToMMo とちとせ研究所は 2022 年度に実施する取組を企画している。ToMMo で分譲可能となっている各種データを用いた試行的解析(データ利活用勉強会)と、まだ分譲に至っていないデータを用いた共同解析(ハッカソン)の実現のための議論を進めている(図 10, 図 11)。

ハッカソンを通じてデータ解析を経て多種多様な機械学習モデルが得られるが、それらを大別すると、「健康状態を判別するモデル」と「行動変容を

促すモデル」に集約される。作成された機械学習モデルはシステム化され、アプリケーションの形で生活者の行動変容を促すことが可能になる。さらに、これら学習済みモデルは運用を通じて再訓練されるが、過学習の恐れもあるため、定期的にコホートデータ等を用いた保守・運用が必要であるため、AI によるデータ駆動型ヘルスケア産業において、既に取得されたヘルスケアデータの重要度は今後ますます増大することが期待される(図 12)。

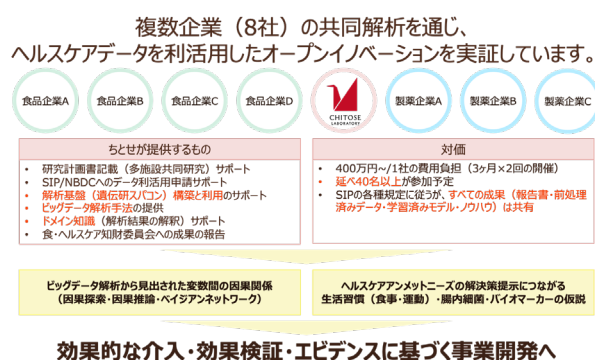


図7. 試行的産業利用実例としての SIP ハッカソン

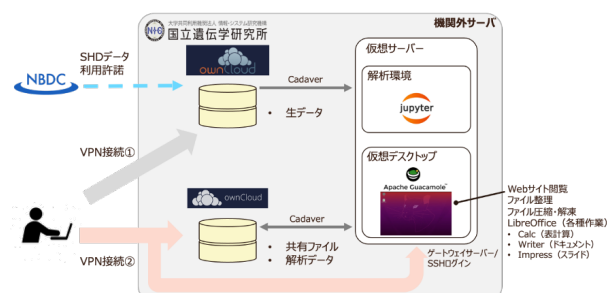


図8. 国立遺伝学研究所が提供するスーパーコンピュータ上に構築したデータ解析プラットフォーム

マイクロバイオームデータ解析を可能にするスパコン解析環境		
	SIP 2B (遺伝研スコン)	SIP 2B-PRISM連携 (ToMMoスコン)
複数拠点からのVPN接続+httpsによる外部インターネット接続	○	制限あり
統計解析+機械学習環境	○	○
スプレッドシートを用いたデータの操作	○	検討中
テーブルデータ作成パイプラインの提供 (Fastqファイルの前処理・マッピング・カウント テーブルの作成)	実施しない	○
クラウドストレージによる任意グループ・個人へのアクセス権限の付与	○	検討中

図9. SIP 事業と SIP-PRISM 連携において構築されたマイクロバイオームデータ解析環境の概要

	PRISM試行的産業利用 データ活用勉強会 ハズオン講座	TOMMO-ちとせ共同研究の枠組みに参画 共同解析（ハッカソン）
利用データと実施内容	過去のTOMMO分譲データ+300人分マイクロバイオームデータを用いた解析ハズオン+試行的解析	TOMMOで取得している任意のデータ（分譲の範囲を問わず、TOMMOが可能なもの：詳細三次元）+300人分のマイクロバイオームデータを用いたハッカソン
実施・契約形態	TOMMOからのデータ分譲	TOMMO-CLの共同研究に参画
参画費用	¥0～分譲費用等を按分（検討中）	2000万円（ちとせ人件費）+TOMMO共同研究費用等を按分（検討中）
解析時期・期間	2022年後半のうち1ヶ月程度（週1回×4回程度を想定）（検討中）	2022年末～2023年初頭のうち2～3ヶ月程度（週1回～の頻度を想定）（検討中）
成果物の取り扱い	基本方針として、TOMMO分譲規定（研究試料及び研究情報の分譲に関する契約書）に従い、TOMMOは権利を主張しない。ただし、一部、教育目的でTOMMOが利用する可能性がある	基本方針として、TOMMOの共同研究の方針に従う。（後日協議といった記載は可能） ・前処理済みデータをTOMMOと参画企業間で共有し、作成した学習済みモデルを参画企業が自己実施できる形を想定（検討中）

図10. 2022 年度実施を計画中の試行的産業利用

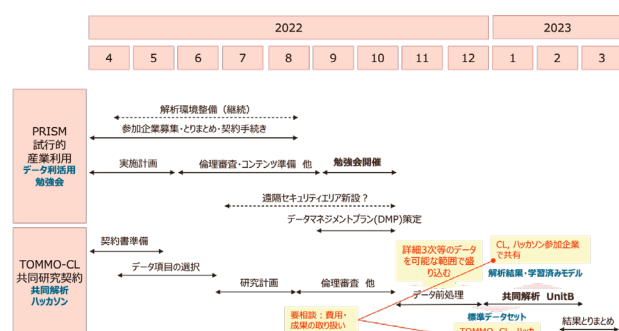


図 11. 2022 年度実施のスケジュール案

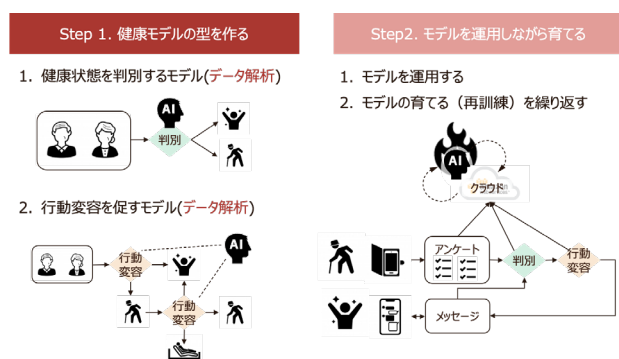


図 12. ハッカソン（データ解析）で得られる成果

D. 考察

アプローチした 40 社を超える企業のうち約 4 割の企業がコホート・バイオバンクの積極的な利活用に意欲を示し、ハッカソンや共同研究などを検討もしくは実施する段階へと進むこととなった(図1)。あらかじめコホート・バイオバンクに対する潜在的な需要が見込まれる企業を選定していたとはいえ、この数値は驚愕すべきものである。これまでの PRISM 事業の中でも、我が国の産業界のコホート・

バイオバンク利活用に対する期待の高さは認められていたが、今回の研究により改めてコホート・バイオバンクに対するニーズの高さが実証されたと言える。

「バイオ戦略 2020」が掲げる「2030 年に世界最先端のバイオエコノミー社会を実現」するためには、産学が一体となった国際競争力を有した研究開発の推進、市場開拓が求められている。産業界とコホート拠点との連携促進はその鍵となるものであり、今回の研究で明らかになった産業界のコホート・バイオバンクに対する高いニーズを無駄なく効率的に研究開発につなげるためには、本 PRISM 事業も取り組んでいる産業界とコホート・バイオバンクを繋ぐインフラの整備は、我が国の産学官が一体となって取り組むべき課題であると考えられる。

E. 結論

産業界とコホート拠点との連携促進を図るために協調可能なバイオデータ等を活用した試行的取組をおこなった。40 社を超える企業にアプローチし、そのうち約 4 割の企業がコホート・バイオバンクを活用したハッカソンや共同研究などを検討する段階へと進むなど我が国の産業界のコホート・バイオバンクに対する高いニーズが改めて実証された。

また、コホート・バイオバンクの利活用を検討している企業に対して、本 PRISM 事業で構築した窓口機能を介して研究開始に必要な研究計画立案や倫理申請手続き等のコーディネーションをおこなった。窓口機能が効率的に機能することで研究開始までの労力や時間が軽減され、産業界によるコホート・バイオバンクの効率的な利活用につながる事が示された。

2022 年度中に産業界とのデータ利活用勉強会やハッカソンを行うことを計画しており、その実施に向けて解析環境の整備等をおこなっている。8 社を超える多くの企業が参加する予定であり、コホート・バイオバンクに対して高いニーズを持つ企業が実際にコホート・バイオバンクの利活用を経験することで、今後より積極的かつ効率的な産学連携研究

開発の実施につながると期待される。

F. 健康危機情報 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