



ライフサイエンスデータベース統合推進事業

統合化推進プログラム

2025年度公募 募集説明会

2024年12月25日（水）

国立研究開発法人 科学技術振興機構（JST）

情報基盤事業部 NBDC事業推進室

本日のプログラム

1. はじめに
2. 統合化推進プログラムについて
3. 研究総括方針
4. 提案・選考に関する注意事項など
5. 質疑応答
6. お問い合わせ先

注意事項

本公募は2025年度予算によるもので、国会における2025年度政府予算の成立を前提として行います。

予算編成の状況によっては、公募の内容を予告なく変更する場合があります。予めご了承ください。

募集説明会にあたって

1. 本日はZOOMウェビナーで行います。
2. 「チャット機能」はご使用いただけません。
3. 「研究総括方針」、「提案・選考に関する注意事項など」の説明の後に、ご質問に対する回答時間を設けますので、ご質問はZOOMの「Q&A」に書き込んでください。説明中に書き込んでいただいてもかまいません。尚、内容によっては、この場で回答できない場合もありますので、ご了承ください。
4. 個別に回答をご希望される方は、メールにてご質問をお送りください。
5. この場でお答えできなかった質問について回答が必要な場合は、改めて、メールにて下記までご質問をお送りください。

ご質問の送り先：nbdc-funding@jst.go.jp

募集説明会の動画とQ&Aの公開

- 質疑応答のパートを除いた募集説明会の動画とプレゼン資料、ならびに本日の質疑応答でご紹介した質問と回答は、近日中にNBDC公募サイトから公開いたします。
- NBDC公募サイト

<https://biosciencedbc.jp/funding/calls/2025.html>

統合化推進プログラムについて

統合化推進プログラムとは

- 統合化推進プログラムは、ライフサイエンスに関わる国内外のデータを統合的に扱うための統合データベースの開発を目的としたプログラム。
- 国際的なデータ共有の推進とオープンサイエンスへの貢献を目指し、国際的に運用されている一次レポジトリや、新たな知識創出を目的とした二次データベースなど、幅広いライフサイエンス分野のさまざまなデータベースの研究開発を支援してきた。

「統合化推進プログラム」採択課題

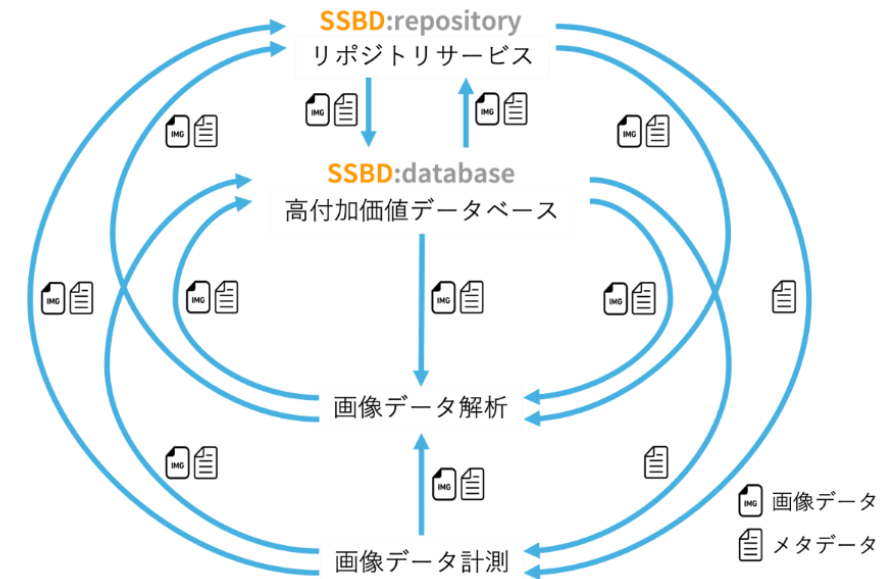
区分	研究開発課題名	研究代表者	所属機関	開発対象データベース	採択年度
本 格 型	バイオイメージングデータのグローバルなデータ共有システムの構築	大浪 修一	理化学研究所	SSBD Database	2022年度
	統合的な転写制御データ基盤の構築	粕川 雄也	理化学研究所	INTRARED (fanta.bio/ChIP-Atlas)	2022年度
	ヒトゲノム・病原体ゲノムと疾患・医薬品をつなぐ統合データベース	金久 實	京都大学	KEGG MEDICUS	2022年度
	異分野融合を志向した糖鎖科学ポータルデータの拡充と品質向上	木下 聖子	創価大学	GlyCosmos Glycoscience Portal	2022年度
	蛋白質構造データバンクのデータ駆動型研究基盤への拡張	栗栖 源嗣	大阪大学	PDBj (Protein Data Bank Japan)	2022年度
	マイクロバイーム研究を先導するハブを目指した微生物統合データベースの特化型開発	森 宙史	国立遺伝学研究所	Microbiome Datahub (開発中)	2022年度
	jPOST prime : コミュニティ連携を基盤とするプロテオームデータベース環境の実現	石濱 泰	京都大学	Japan Proteome Standard Repository/Database (jPOST)	2023年度
	次世代低分子マスペクトルデータベース シン・マスバンクの構築	松田 史生	大阪大学	Shin-MassBank	2023年度
育 成 型	非モデル植物のための遺伝子ネットワーク情報活用基盤	大林 武	東北大学	ATTED-II	2023年度
	日本人塩基配列情報の公開可能なゲノム・オミクス情報基盤による双方向型研究教育データベース開発と国際連携	長崎 正朗	九州大学	Japanese Open Genome Omics Platform (JoGo) (開発中)	2023年度
	空間オミクスデータ解析用データベースの開発	VANDENBON Alexis	京都大学	DeepSpaceDB	2023年度
	AI駆動型データキュレーションによる持続可能な中分子相互作用統合データベースの開発	池田 和由	理化学研究所	MIIDB-AI (開発中)	2024年度
	細胞レベルの機能・表現型と遺伝子発現を関連付ける「Cell IO」データベースの開発	尾崎 遼	筑波大学	Cell IO (開発中)	2024年度
	創発的再解析のためのメタボローム統合データベース	早川 英介	理化学研究所	integMET (開発中)	2024年度

統合化推進プログラムのデータベース構築例（本格型）（1）

○ 生物画像データベース SSBD [理研 大浪修一] **SSBD:database** (<https://ssbd.riken.jp/database/>)

バイオイメージングデータと生命動態定量データの統合データベース。日本の全ての生命科学研究者が利活用し、かつ日米欧などの国際連携に基づく国際的データ共有基盤として再構築する。情報科学的な画像解析手法の開発・共有促進プラットフォームの構築、持続的な運用と発展を可能にするためのシステム開発を実施する。

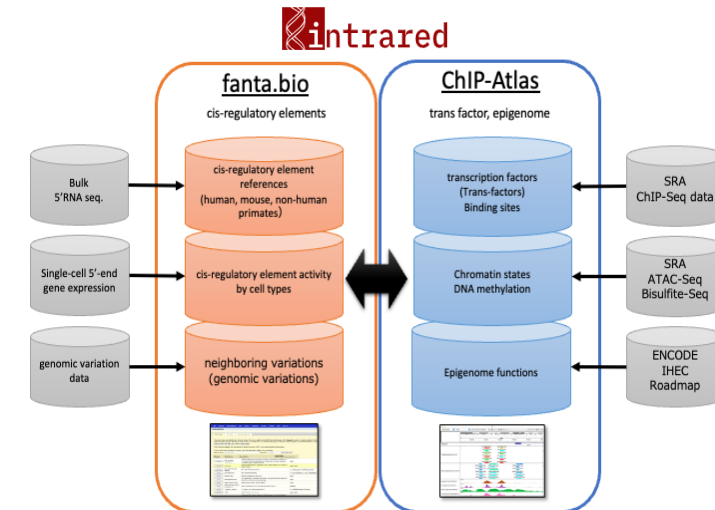
- ・ **SSBD:repository** : 生物画像データの公共リポジトリ。
- ・ **SSBD:database** : メタデータリッチな高付加価値データベース。



○ 統合的転写制御解析基盤 Intrared (理研 粕川 雄也) **Intrared** (<https://www.intrared.org/>)

シスエレメントデータベースfanta.bioとエピゲノミクスデータベースChIP-Atlasを統合し、統合的転写制御解析基盤を構築する。

- ・ **fanta.bio** : 遺伝子発現調節に関わるゲノム上のシス転写調節配列（シスエレメント）と細胞の種類・状態ごとの活性状態を収録するデータベース。ヒト、マウスのシスエレメント情報と疾患関連多型や変異情報等を収載。
- ・ **ChIP-Atlas** : ChIP-seq、ATAC-seq、DNase-seq、Bisulfite-seqのデータをそれぞれ統一プロトコルで再解析して抗原別・生物種別・細胞種別に整理し、Hi-C, ChromHMM, GWAS SNPs、eQTLsデータなどの様々なエピゲノミクスを統合したデータベース。



統合化推進プログラムのデータベース構築例（本格型）（2）

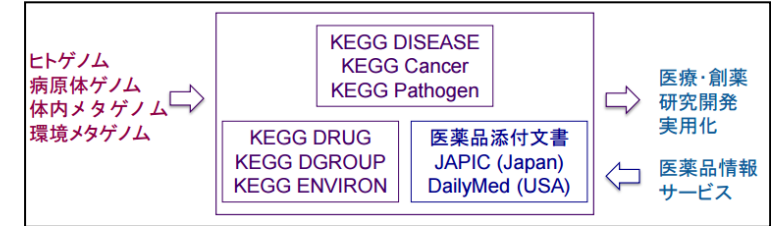
○ ゲノム・疾患・医薬品の統合データベース KEGG MEDICUS [京都大学 金久 寛]



(<https://www.kegg.jp/kegg/medicus/>)

個々の遺伝子や複数の遺伝子から構成された機能モジュール、遺伝子、タンパク質、環境因子、医薬品等から構成された相互作用ユニットに関する知識をデータベース化し、ゲノム情報を有効利用するためのリソース。

- ・ 遺伝子、タンパク質、また代謝やシグナル伝達などの分子間ネットワークを統合したデータベースKEGGで、ヒトゲノム、病原体ゲノムなどのシーケンス解読と有効利用を促進する統合リソースKEGG MEDICUSを開発し提供。
- ・ KEGG MEDICUSの月間平均利用者数は約280万人、利用件数は600万件を超える。Googleで医薬品名を検索すると上位にヒットし、一般市民からも広く利用されている。



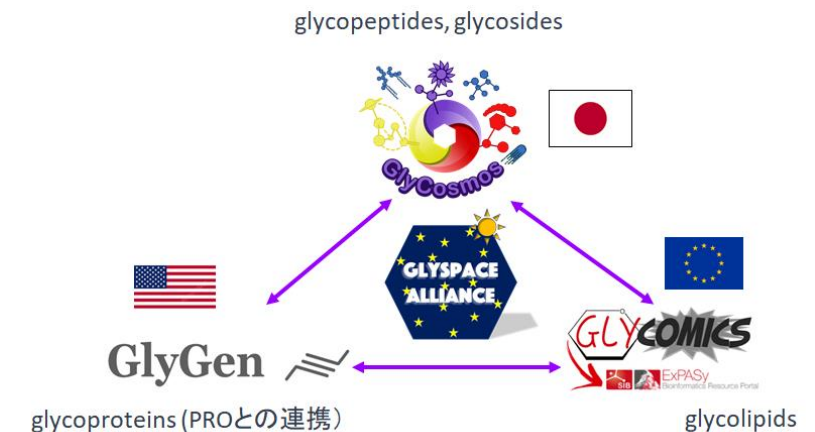
KEGG MEDICUSの構成。日米の医薬品添付文書、疾患情報、医薬品情報を整備し、ゲノム、メタゲノム情報と統合して提供している。

○ 糖鎖情報統合データベース GlyCosmos Glycoscience Portal [創価大学 木下 聖子]

(<https://glycosmos.org/>)

国際コンソシアムGlySpace Allianceと連携し、糖鎖構造のIDや構造表記、フォーマット等の共通化を図る。

- ・ リポジトリに登録された様々な糖鎖関連データをキュレーションして公開する仕組みを構築し、分野を超えた糖鎖情報の活用を促進して、糖鎖科学と構造生物学、微生物、植物、プロテオミクスなどとの融合を推進する。
- ・ これまで構築してきた糖鎖構造、複合糖鎖、糖鎖分析データなどレポジトリに加え、レクチンマイクロアレイデータや微生物関連糖鎖情報および糖鎖関連パスウェイデータなどにもレポジトリを拡張する。



統合化推進プログラムのデータベース構築例（本格型）（3）

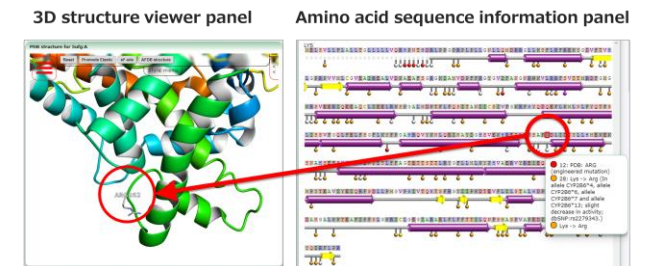
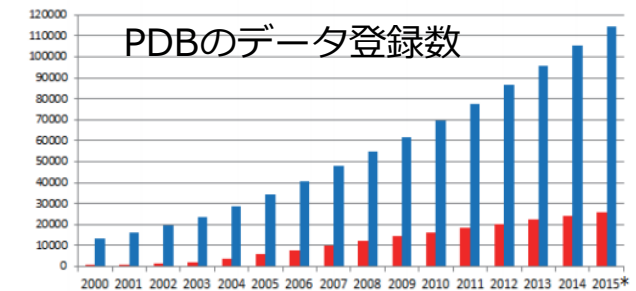
○ タンパク質構造データバンク PDBj [大阪大学 栗栖 源嗣]



(<https://pdbj.org/>)

PDB（タンパク質構造データバンク）とBMRB（NMR実験情報データバンク）を構築・公開し、日米欧3極で構成するwwPDBの主要メンバーとしてアジア地区からのタンパク質立体構造の登録を処理。独自のアノテーションやツールを提供し、構造生物学データのさらなる高度化を図る。

- ・世界的な生命科学研究基盤データベースとして認知されており、2022年、Global Biodata Coalition (GBC)のGlobal Core Biodata Resourcesに認定。
- ・世界全体で登録された約20万件のデータのうち約24%をPDBjが登録。
- ・全ての実験手法でデータ品質の検証レポートを確立しRDF化して利用者に提供。
- ・近年、急増している予測構造を含む構造にも対応。
- ・日本人ゲノムバリエーション情報をPDB構造上にマップし、生物学・化学の視点からの検索結果を表示する「PDB統合利用ポータル」を2024年10月に公開。
- ・構造ベース創薬（SBDD）など応用分野でも広く利用されている。



○ マイクロバーム統合データベース Microbiome Datahub (国立遺伝研 森 宙史)

(<https://mdatahub.org/>)（開発中）

これまで構築してきた微生物ゲノム統合データベース MicrobeDB.jpを発展させ、国際的に隆盛をきわめる微生物叢（マイクロバイオーム）研究に特化した国際的なデータハブ Microbiome Datahubとして再構築する。微生物の表現型やメタゲノム由来のゲノム（Metagenome Assembled Genome; MAG）などの新しいデータを統合化し、マイクロバイオーム関連のデータ検索・解析機能を大幅に強化して、自由度の高い検索・データ取得を実現する。

統合化推進プログラムのデータベース構築例（本格型）（4）

○ プロテオーム統合データベース jPOST [京都大学 石濱 泰]



(<https://jpostdb.org/>)

国内外に散在するプロテオーム情報を標準化・統合・一元管理し、多彩な生物種・翻訳後修飾・絶対発現量も含めた統合プロテオームデータベースを開発。

- ・プロテオーム国際コンソーシアム（ProteomeXchange）のアジア初メンバー。ProteomeXchangeではデータIDを共通化、各サイトに登録されているデータ検索を可能にした。
- ・独自に開発した品質評価指標UniScoreを用いた再解析によりデータセットを跨ぐ統合解析を可能にした。
- ・2022年、jPOSTを含むProteomeXchangeがGlobal Biodata Coalition (GBC)のGlobal Core Biodata Resourcesに認定、国際的に重要なバイオデータベースの一つとして認知された。
- ・アジアや欧米からも多くのデータが登録され、国内外の研究者により広くデータ利用がされている。

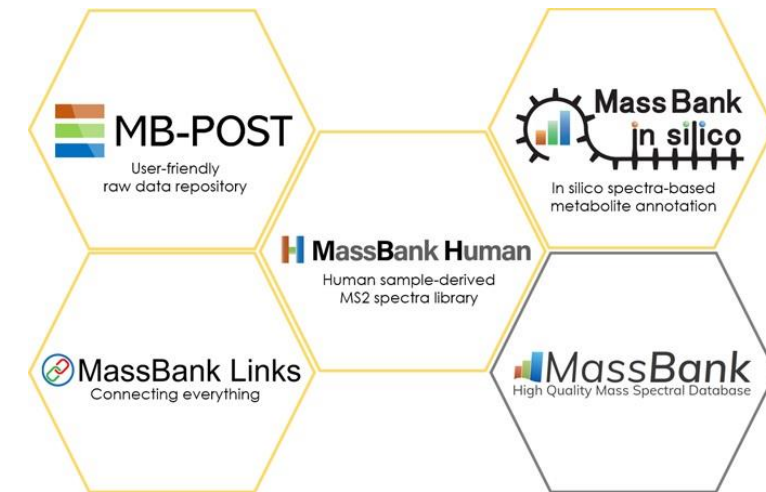


○ メタボローム解析パイプライン Shin-MassBank [大阪大学 松田 史生]

(<https://shin.massbank.jp/>)（開発中）

生体由来スペクトルを集積していくための解析パイプライン Shin-MassBankを構築し、高品質な実測マススペクトルをMassBankに保存していく。ユーザー主導の開発体制で、生命科学研究者などのユーザーニーズを満たす次世代データベースを構築する。

- ・MB-POST：質量分析生データリポジトリ
- ・MassBank human：ヒトサンプル由来のスペクトラムDB
- ・MassBank *in silico*：予測マススペクトルを収載
- ・MassBank Links



統合化推進プログラム 2023年度 採択課題（育成型）

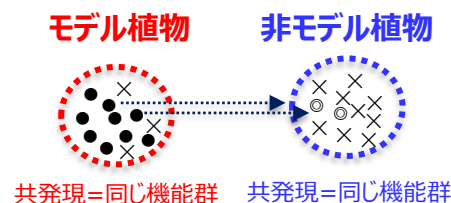
植物遺伝子の共発現DB

ATTED-II

(<https://atted.jp/>)

東北大学 大林武

植物遺伝子の共発現データベース ATTED-IIをベースに、モデル植物の遺伝子共発現情報を充実させ、モデル植物で得られている情報を非モデル植物へと橋渡しする仕組みを構築する。これにより、モデル植物の遺伝子共発現ネットワークから得られる知見を非モデル植物の研究に活用することが可能になる。また、遺伝子が共発現する条件や環境要因とその遺伝子が担う生理学的な機能やその制御関係を考察するための解析ツールを開発して実装する。



<2024年10月に公開したver.12.0の新機能>

- 新たにコムギとオオムギのデータを追加。
- 着目する共発現遺伝子ペアに関連する実験条件や環境要因を特定するCoexViewerを新たに実装。
- 特定の条件や環境下でどのような遺伝子が共発現しているかを示す主成分分析ツールPC Viewを新たに実装。

日本人ゲノムの構造多型DB

JoGo Platform

(開発中)

九州大学 長崎正朗

塩基配列レベルで公開可能な日本人由来の約100サンプルを長鎖シーケンスによって決定した全ゲノム配列とそこから明らかになった構造多型情報、および該当サンプルのマルチオミクス情報を含むデータベースを構築する。本データベースは、日本人における構造多型と希少疾患や多因子疾患との関連性を理解する上で重要な情報基盤となる。



空間トランスクリプトミクスDB

DeepSpaceDB

(<https://www.deepspacedb.com>)

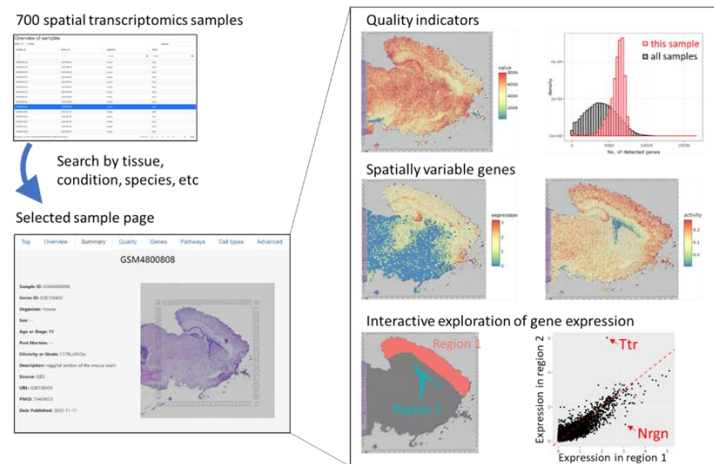
京都大学 Vandenberg Alexis

さまざまな組織の空間トランスクリプトームデータを、誰もが簡単に高度なレベルで解析できるようにするデータベース。空間トランスクリプトミクス解析の主要プラットフォームである10x Visiumで取られたヒトおよびマウスの公開データをほぼ網羅的に収集して掲載。複雑な内部構造を有する組織を視覚的に理解したり、腫瘍サンプルの微小環境を理解することが可能になる。



<2024年8月に公開したver.1の概要>

- NCBIなどで公開されている10xVisiumデータを収集、数値化とQCを実施し品質基準を満たす1,000件以上のヒトとマウスのデータの組織や疾患状態等のメタ情報を整備して掲載。
- 任意の遺伝子の空間的発現分布の表示。
- 発現している遺伝子から予測された空間的に発現が変動しているパスウェイの表示。
- 発現している遺伝子から予測された各スポットにある細胞の種類の表示。
- 任意に選択したエリア間での遺伝子発現比較ツールを実装。
- 微小環境を遺伝子発現パターンの類似性に基づいてグループ化。



DeepSpaceDBインターフェイスでの表示例

(特定遺伝子の発現状態や選択領域で特徴的な遺伝子を可視化)

統合化推進プログラム 2024年度 採択課題（育成型）

AI駆動型データキュレーションによる持続可能な中分子相互作用統合データベースの開発

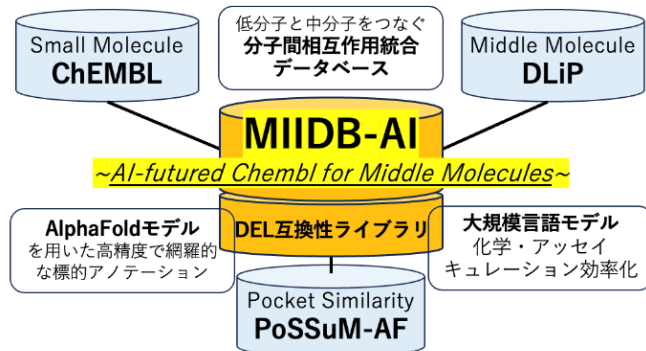


MIIDB-AI (開発中)

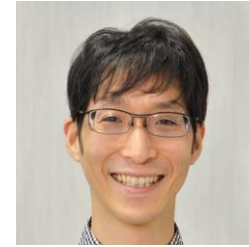
理化学研究所 池田 和由

<研究開発の概要>

創薬が困難な標的分子の情報を収載することを目的として、中分子（ペプチド、非ペプチド、核酸）の相互作用データベースMIIDB-AIを構築する。本データベースでは、リガンド結合部位類似性データに基づき中分子の標的相互作用部位を特定し、AI技術を使って高精度に標的とリガンド間の相互作用を予測することができる。中分子の知識基盤データベースによって新規の中分子医薬品候補を効率良く見つけ出せるようになり、次世代創薬研究への貢献が期待できる。



細胞レベルの機能・表現型と遺伝子発現を関連付ける「Cell IO」データベースの開発

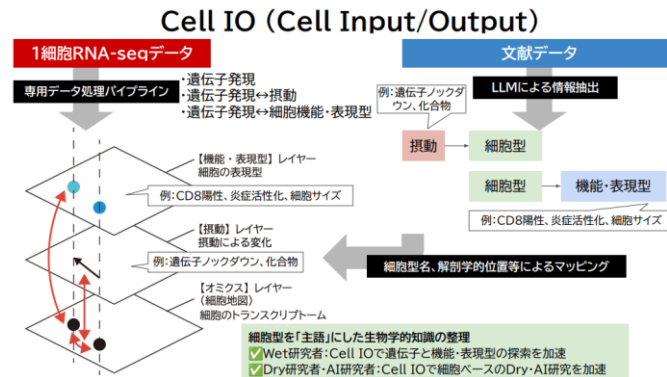


Cell IO (開発中)

筑波大学 尾崎 遼

<研究開発の概要>

細胞の機能・表現型を整理したデータベース「Cell IO」を構築する。ヒトやマウスなどの1細胞トランスクリプトームデータと紐づく細胞の機能・表現型と摂動情報を整理する。文献から細胞の由来、機能・表現型、摂動の種類、生物種等の情報を抽出し、検索やデータ統合を容易にする。遺伝子発現を細胞の機能・表現型として解釈可能にするとともに、オミクスデータから細胞の機能・表現型を予測する研究や所望の機能・表現型を持つ細胞をデザインする研究などに貢献する。



創発的再解析のためのメタローム統合データベース

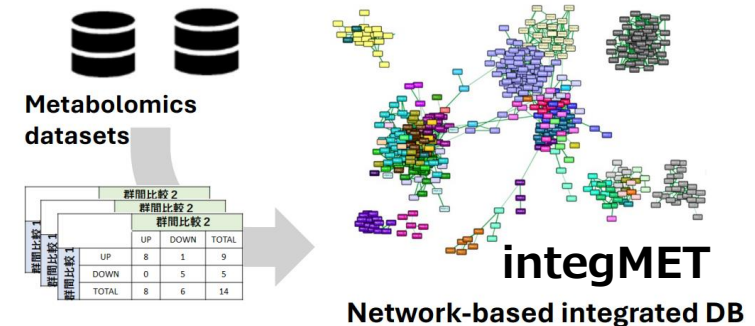


integMET (開発中)

理化学研究所 早川 英介

<研究開発の概要>

代謝物の変動傾向とメタデータの類似度からメタロームデータを知識ネットワークとして整備したデータベースintegMETを構築する。ネットワークとして体系化したメタロミクスデータの大規模データ再解析を可能にすることで、代謝と生命現象との関係を俯瞰的に理解できるようにする。さらに、研究情報を統合したメタローム知識グラフ構造をデータサイエンスやAI研究に応用するなど、分野横断的な新しい研究・知識の創出につながるデータベースを目指す。



統合化推進プログラム
研究総括方針

九州大学大学院 医学研究院 教授

伊藤隆司

統合化推進プログラムの公募開始にあたって

伊藤 隆司（九州大学 医学研究院）

統合化推進プログラム（DICP）とは

統合化推進プログラムは、ライフサイエンスに関わる国内外のデータを統合的に扱うためのデータベース（統合データベース）開発を目的としたプログラムです。

具体的な研究開発は、研究データの収集・整理、標準化・公開・共有・品質管理、運用等で、他のデータベースとの連携・統合化や、収録データを円滑に利用するために必要なインターフェース・ツールの開発等も含まれます。

本プログラムは、これらの研究開発を通じて**公共データ利活用のための情報環境整備**を行うとともに、利用者の知識発見や課題解決への寄与および国際的なオープンサイエンスへの貢献を目指します。

 **健全なバイオデータサイエンスの発展こそが究極の目的**

統合化推進プログラム（DICP）の歴史

- 2011年に第1期が開始
 - 過去3期で計31課題を支援 (<https://biosciencedbc.jp/funding/project/>)
- 2022年に第4期が開始
 - 広範なユーザーの知識発見の支援
 - 国際的プレゼンス
 - 新しい動向への対応
 - 「つなぐ」から「使う」へ

研究アドバイザー

鎌田 真由美	北里大学 未来工学部 教授
坂井 寛章	農業・食品産業技術総合研究機構 高度分析研究センター ユニット長
清水 佳奈	早稲田大学 理工学術院 教授
瀬々 潤	(株) ヒューマノーム研究所 代表取締役社長
馬場 健史	九州大学 生体防御医学研究所 教授
山本 一夫	お茶の水女子大学 ヒューマンライフサイエンス研究所 客員教授
吉田 哲郎	アクセリード (株) 経営企画部 シニアディレクター

(2024年9月現在；五十音順に掲載)

DICP第4期で目指して頂きたいこと

- 当該コミュニティに支持され愛用されるDB
- コミュニティ外の広範なユーザーの知識発見の支援
- 国際的なプレゼンス：協調・突出
- 課題間の相乗効果：データの統合と技術の共有
- 新しい動向を先取りした対応

DICP第4期で目指したいこと

ライフステージに応じたDB支援の複線化

- 試験的支援による発掘・育成 **DICP育成型**
- 本格的支援による成長・確立 **DICP本格型**
- 別枠の支援による維持・継続 **?**
- 現役引退データベースの保全 **生命科学系データベース
アーカイブ**

統合化推進プログラム（DICP）の歴史

- 2011年に第1期が開始
 - 過去3期で計31課題を支援 (<https://biosciencedbc.jp/funding/project/>)
- 2022年に第4期が開始
 - 広範なユーザーの知識発見の支援
 - 国際的プレゼンス
 - 新しい動向への対応
 - 「つなぐ」から「使う」
 - 2023年より**育成型**を開始
(DICPの新しい挑戦)

研究アドバイザー

鎌田 真由美	北里大学 未来工学部 教授
坂井 寛章	農業・食品産業技術総合研究機構 高度分析研究センター ユニット長
清水 佳奈	早稲田大学 理工学術院 教授
瀬々 潤	(株) ヒューマノーム研究所 代表取締役社長
馬場 健史	九州大学 生体防御医学研究所 教授
山本 一夫	お茶の水女子大学 ヒューマンライフサイエンス研究所 客員教授
吉田 哲郎	アクセリード (株) 経営企画部 シニアディレクター

(2024年9月現在；五十音順に掲載)

今回の募集：育成型の3期生

新たなデータベースの構築を目指す萌芽的な研究開発

技術動向や研究ニーズへの対応を目指し、**未整備分野**の萌芽的な統合データベースの構築や、**斬新な切り口**によるデータ統合など、独創的な発想に基づいた知識発見・課題解決の支援に資する先駆的なデータベースを研究開発する提案を募集します。

新たにデータベースを構築しようとしている、既にデータベースを構築しているが未公開の状態にある、データベースを公開していて将来的に国際的な地位確立を目指している場合などを対象とします。

いずれは国際基準のデータ基盤へ発展していくことを期待しています。

育成型に求めること

課題全般に求めること

- 広範なユーザーの知識発見の支援
- 国際的プレゼンス
- 新しい動向への対応
- 「つなぐ」から「使う」へ

育成型に求めること

- 構想に新規性・独自性・挑戦性・将来性が感じられること
- 実績は問わないが、実現可能性が感じられること
 - DB構築の戦略に不安を感じさせない陣容
- 既存DBの新展開の提案に、独自性・挑戦性・将来性が感じられること
 - 従来の延長線上に留まらない飛躍・新機軸

現在支援中の育成型課題（1期生）

植物遺伝子の共発現DB

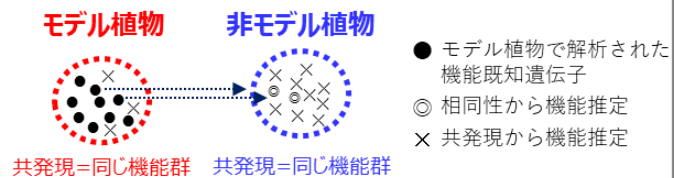
ATTED-II

(<https://atted.jp/>)

東北大学 大林武



植物遺伝子の共発現データベース ATTED-IIをベースに、モデル植物の遺伝子共発現情報を充実させ、モデル植物で得られている情報を**非モデル植物**へと橋渡しする仕組みを構築する。これにより、モデル植物の遺伝子共発現ネットワークから得られる知見を非モデル植物の研究に活用することが可能になる。また、遺伝子が共発現する条件や環境要因とその遺伝子が担う生理学的な機能やその制御関係を考察するための解析ツールを開発して実装する。



<2024年10月に公開したver.12.0の新機能>

- 新たにコムギとオオムギのデータを追加。
- 着目する共発現遺伝子ペアに関連する実験条件や環境要因を特定するCoexViewerを新たに実装。
- 特定の条件や環境下でどのような遺伝子が共発現しているかを示す主成分分析ツールPC Viewを新たに実装。

日本人ゲノムの構造多型DB

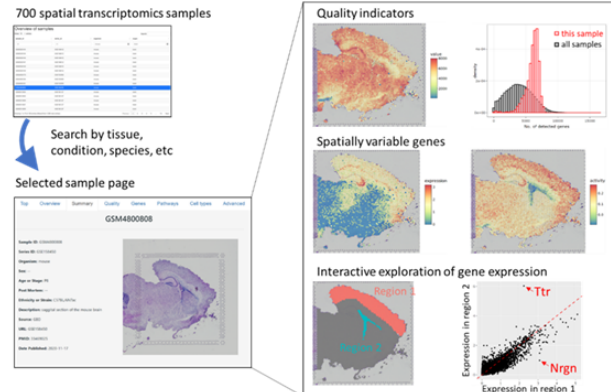
JoGo Platform

(開発中)

九州大学 長崎正朗



塩基配列レベルで公開可能な日本人由来の約100サンプルを**長鎖シーケンス**によって決定した全ゲノム配列とそこから明らかになった構造多型情報、および該当サンプルのマルチオミクス情報を含むデータベースを構築する。本データベースは、日本人における構造多型と希少疾患や多因子疾患との関連性を理解する上で重要な情報基盤となる。



DeepSpaceDBインターフェイスでの表示例

(特定遺伝子の発現状態や選択領域で特徴的な遺伝子を可視化)

空間トランスクリプトミクスDB

DeepSpaceDB

(<https://www.deepspacedb.com>)

京都大学 Vandenberg Alexis



さまざまな組織の**空間トランスクリプトーム**データを、誰もが簡単に高度なレベルで解析できるようにするデータベース。空間トランスクリプトミクス解析の主要プラットフォームである10x Visiumで取られたヒトおよびマウスの公開データをほぼ網羅的に収集して収載。複雑な内部構造を有する組織を視覚的に理解したり、腫瘍サンプルの微小環境を理解することが可能になる。

<2024年8月に公開したver.1の概要>

- NCBIなどで公開されている10xVisiumデータを収集、数値化とQCを実施し品質基準を満たす1,000件以上のヒトとマウスのデータの組織や疾患状態等のメタ情報を整備して収載。
- 任意の遺伝子の空間的発現分布の表示。
- 発現している遺伝子から予測された空間的に発現が変動しているパスウェイの表示。
- 発現している遺伝子から予測された各スポットにある細胞の種類の表示。
- 任意に選択したエリア間での遺伝子発現比較ツールを実装。
- 微小環境を遺伝子発現パターンの類似性に基づいてグループ化。

現在支援中の育成型課題（2期生）

AI駆動型データキュレーションによる持続可能な中分子相互作用統合データベースの開発

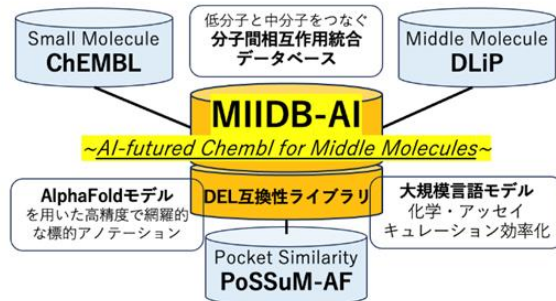


MIIDB-AI（開発中）

理化学研究所 池田 和由

<研究開発の概要>

創薬が困難な標的分子の情報を収載することを目的として、**中分子**（ペプチド、非ペプチド、核酸）の相互作用データベースMIIDB-AIを構築する。本データベースでは、リガンド結合部位類似性データに基づき中分子の標的相互作用部位を特定し、**AI技術**を使って高精度に標的とリガンド間の相互作用を予測することができる。中分子の知識基盤データベースによって新規の中分子医薬品候補を効率良く見つけ出せるようになり、次世代創薬研究への貢献が期待できる。



細胞レベルの機能・表現型と遺伝子発現を関連付ける「Cell IO」データベースの開発

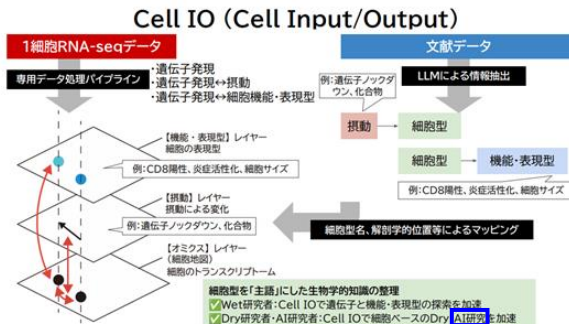


Cell IO（開発中）

筑波大学 尾崎 遼

<研究開発の概要>

細胞の機能・表現型を整理したデータベース「Cell IO」を構築する。ヒトやマウスなどの**1細胞トランスクリプトーム**データと紐づく細胞の機能・表現型と摂動情報を整理する。文献から細胞の由来、機能・表現型、摂動の種類、生物種等の情報を抽出し、検索やデータ統合を容易にする。遺伝子発現を細胞の機能・表現型として解釈可能にするとともに、オミクスデータから細胞の機能・表現型を予測する研究や所望の機能・表現型を持つ細胞をデザインする研究などに貢献する。



創発的再解析のためのメタボローム統合データベース

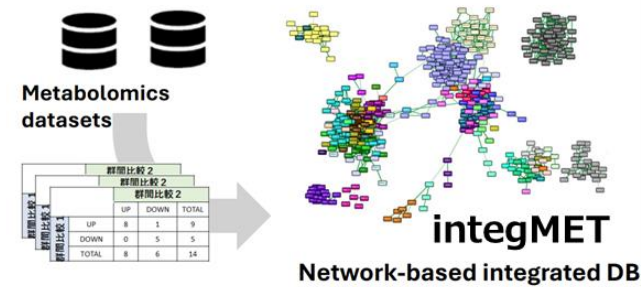


integMET（開発中）

理化学研究所 早川 英介

<研究開発の概要>

代謝物の変動傾向とメタデータの類似度から**メタボロームデータ**を知識ネットワークとして整備したデータベースintegMETを構築する。ネットワークとして体系化したメタボロミクスデータの大規模データ再解析を可能にすることで、代謝と生命現象との関係を俯瞰的に理解できるようにする。さらに、研究情報を統合したメタボローム知識グラフ構造をデータサイエンスや**AI研究**に応用するなど、分野横断的な新しい研究・知識の創出につながるデータベースを目指す。



育成型に求めること

課題全般に求めること

- 広範なユーザーの知識発見の支援
- 国際的プレゼンス
- 新しい動向への対応
 - AIを利用したDB構築の効率化
 - AIによる利用を意識したDB設計
- 「つなぐ」から「使う」へ

育成型に求めること

- 構想に新規性・独自性・挑戦性・将来性が感じられること
- 実績は問わないが、実現可能性が感じられること
 - DB構築の戦略に不安を感じさせない陣容
- 既存DBの新展開の提案に、独自性・挑戦性・将来性が感じられること
 - 従来の延長線上に留まらない飛躍・新機軸

バイオデータサイエンスの流れ

- 実験技術・研究手法の急速な発展
- 研究スタイルの変化
- データ(ベース)の量的・質的変容
- データ(ベース)駆動型研究への期待
- 人工知能の急速な進歩が期待を更に加速

Nobel prize

2002 MS, NMR

2006 RNAi

2007 KO mouse

2008 GFP

2014 Super resolution microscopy

2017 Cryo-EM

2020 CRISPR/Cas

2024 AlphaFold2/protein design

Progress in science depends on new techniques, new discoveries, and new ideas, probably in that order. Sydney Brenner

Highly accurate protein structure prediction with AlphaFold


<https://doi.org/10.1038/s41586-021-03819-2>

Received: 11 May 2021

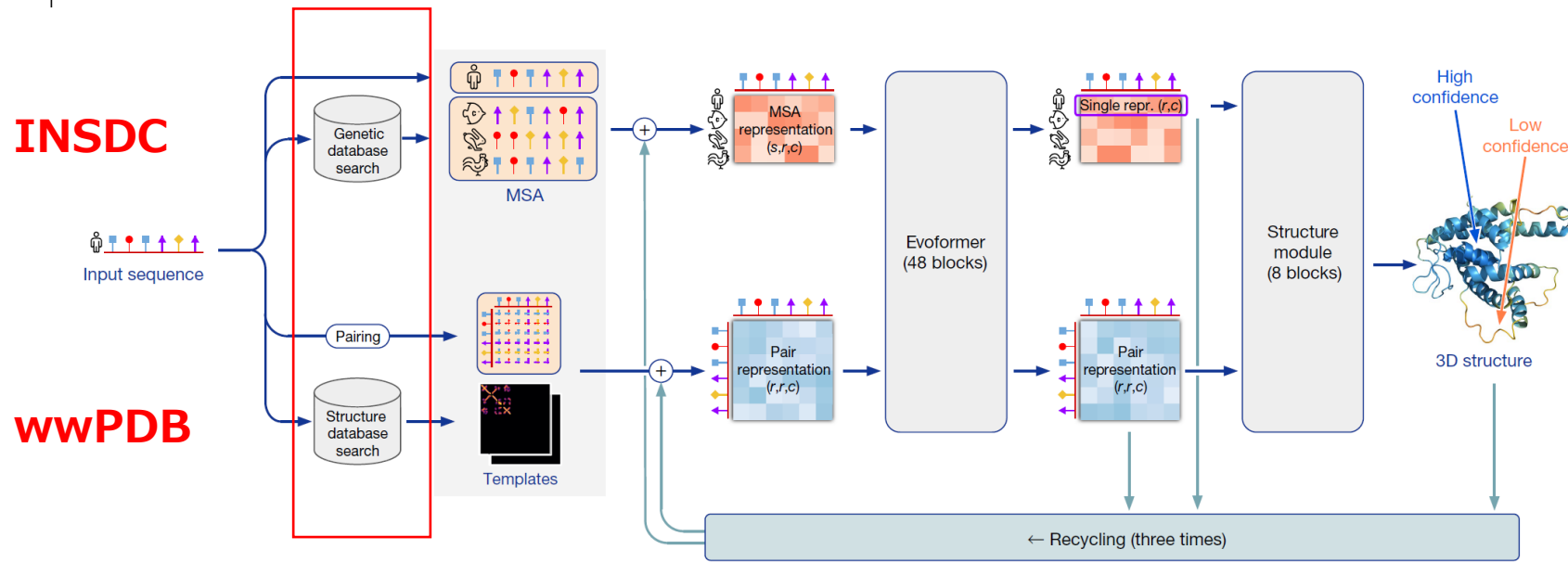
Accepted: 12 July 2021

Published online: 15 July 2021

Open access

 Check for updates

John Jumper^{1,4}✉, **Richard Evans^{1,4}**, **Alexander Pritzel^{1,4}**, **Tim Green^{1,4}**, **Michael Figurnov^{1,4}**, **Olaf Ronneberger^{1,4}**, **Kathryn Tunyasuvunakool^{1,4}**, **Russ Bates^{1,4}**, **Augustin Židek^{1,4}**, **Anna Potapenko^{1,4}**, **Alex Bridgland^{1,4}**, **Clemens Meyer^{1,4}**, **Simon A. A. Kohl^{1,4}**, **Andrew J. Ballard^{1,4}**, **Andrew Cowie^{1,4}**, **Bernardino Romera-Paredes^{1,4}**, **Stanislav Nikolov^{1,4}**, **Rishub Jain^{1,4}**, **Jonas Adler¹**, **Trevor Back¹**, **Stig Petersen¹**, **David Reiman¹**, **Ellen Clancy¹**, **Michal Zielinski¹**, **Martin Steinegger^{2,3}**, **Michalina Pacholska¹**, **Tamas Berghammer¹**, **Sebastian Bodenstein¹**, **David Silver¹**, **Oriol Vinyals¹**, **Andrew W. Senior¹**, **Koray Kavukcuoglu¹**, **Pushmeet Kohli¹** & **Demis Hassabis^{1,4}**✉





Scientific Background to the Nobel Prize in Chemistry 2024

COMPUTATIONAL PROTEIN DESIGN AND
PROTEIN STRUCTURE PREDICTION

The Nobel Committee for Chemistry

Summary and outlook

We are now at a stage where both the structural design and prediction problems are largely solved. The implications of this are far-reaching.

The AlphaFold2 team immediately created large databases of predicted protein structures, first for the human proteome⁴⁴ and then for the majority of sequences (> 200 million) available in the UniProt (Universal Protein Resource) database.⁴⁵ This means that almost overnight, we got access to orders of magnitude more structural information. Likewise, the protein design field has reached a stage where some of the most exciting areas of research are biomedical applications, such as vaccines and protein-based inhibitors,⁴⁶ and applications in synthetic biology.⁴⁷

The progress described above would not have been possible, of course, without the efforts from structural biologists in providing all the experimentally determined structures that have gone into the Protein Data Bank. A number of these spectacular protein structures have also been recognized with Nobel Prizes in Chemistry over the years. These data, resulting from decades of research in protein structure determination, have laid the foundation for the decisive breakthroughs in protein design and structure prediction by this year's Laureates.

In summary, the achievements of **David Baker**, **Demis Hassabis** and **John Jumper** in the fields of computational protein design and protein structure prediction are truly profound. Their work has opened up a new era of biochemical and biological research, where we can now predict and design protein structures in ways that had not been possible before. Hence, a long-standing goal has finally been met, and the impact of this will have far-reaching consequences.

Johan Åqvist

Professor of Theoretical Chemistry

Member of the Royal Swedish Academy of Sciences

Member of the Nobel Committee for Chemistry

バイオデータサイエンスの流れ

- データ(ベース)駆動型研究の時代へ
 - 複数領域のデータを統合する必要性の高まり
 - データ統合による知識発見プロセスの革新
- 👉 AI時代におけるバイオDBのあるべき姿とは？
AI時代のDBの使われ方とは？
AI時代に相応しいDBの作り方とは？

バイオデータサイエンス時代のDB構築

- DBはバイオサイエンスに必須の基盤
AIの力を発揮させるには良質のDBが重要
- DB構築者はバイオサイエンスのエッセンシャルワーカー
DB構築者にとってのAIはIntelligent Assistant
- DB構築の自動化？ (Automated Research Workflow?)
AIがバイオサイエンスのエッセンシャルワーカー？
- AIによるDB設計？
AIの方が新規性の高い研究を提案可能？ (👉)

nature > news > article

NEWS | 20 September 2024

Do AI models produce more original ideas than researchers?

The concepts were judged by reviewers. They were not told who or what had created them.

By Gemma Conroy

[Twitter](#) [Facebook](#) [Email](#)



Can LLMs Generate Novel Research Ideas?

A Large-Scale Human Study with 100+ NLP Researchers

Chenglei Si, Diyi Yang, Tatsunori Hashimoto
Stanford University
{clsi, diyi, thashim}@stanford.edu

Abstract

Recent advancements in large language models (LLMs) have sparked optimism about their potential to accelerate scientific discovery, with a growing number of works proposing research agents that autonomously generate and validate new ideas. Despite this, no evaluations have shown that LLM systems can take the very first step of producing novel, expert-level ideas, let alone perform the entire research process. We address this by establishing an experimental design that evaluates research idea generation while controlling for confounders and performs the first head-to-head comparison between expert NLP researchers and an LLM ideation agent. By recruiting over 100 NLP researchers to write novel ideas and blind reviews of both LLM and human ideas, we obtain the first statistically significant conclusion on current LLM capabilities for research ideation: **we find LLM-generated ideas are judged as more novel ($p < 0.05$) than human expert ideas while being judged slightly weaker on feasibility.** Studying our agent baselines closely, we identify open problems in building and evaluating research agents, including failures of LLM self-evaluation and their lack of diversity in generation. Finally, we acknowledge that human judgements of novelty can be difficult, even by experts, and propose an end-to-end study design which recruits researchers to execute these ideas into full projects, enabling us to study whether these novelty and feasibility judgements result in meaningful differences in research outcome.¹

1 Introduction

The rapid improvement of LLMs, especially in capabilities like knowledge and reasoning, has enabled many new applications in scientific tasks, such as solving challenging mathematical problems (Trinh et al., 2024), assisting scientists in writing proofs (Collins et al., 2024), retrieving related works (Ajith et al., 2024, Press et al., 2024), generating code to solve analytical or computational tasks (Huang et al., 2024, Tian et al., 2024), and discovering patterns in large text corpora (Lam et al., 2024, Zhong et al., 2023). While these are useful applications that can potentially increase the productivity of researchers, it remains an open question whether LLMs can take on the more creative and challenging parts of the research process.

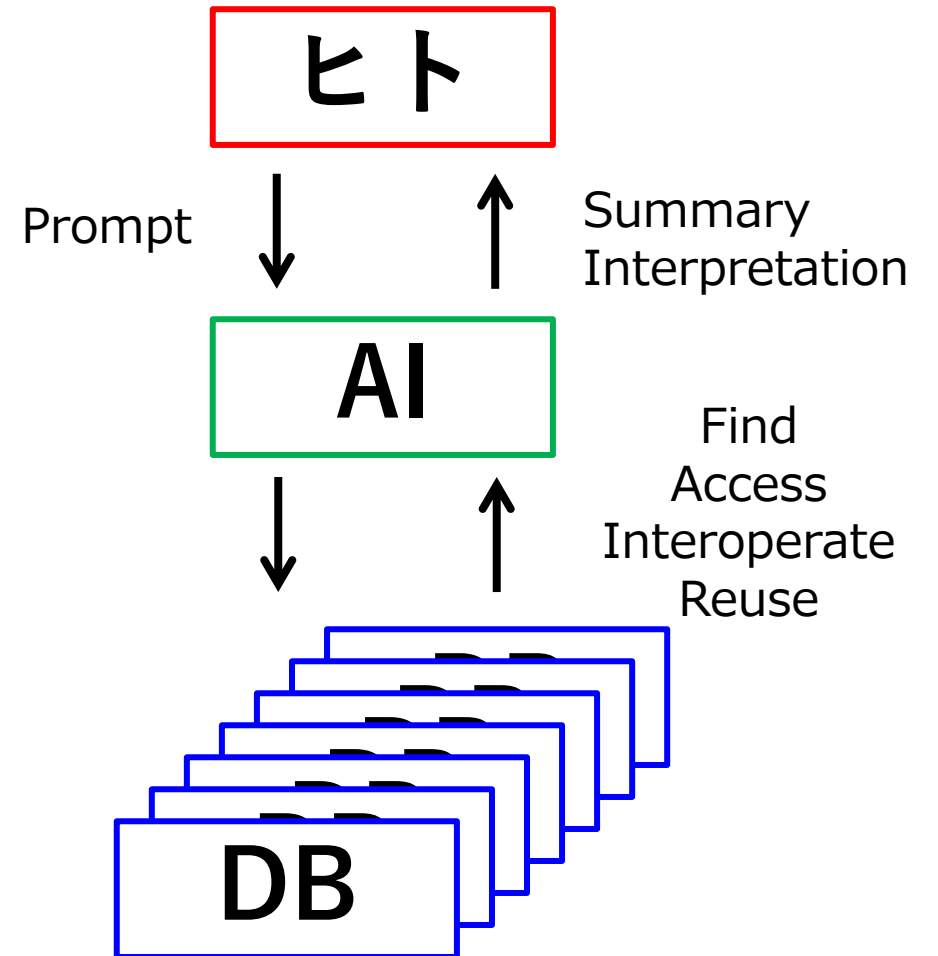
arXiv:2409.04109v1 [cs.CL] 6 Sep 2024

<https://doi.org/10.1038/d41586-024-03070-5>

<https://arxiv.org/abs/2409.04109>

バイオデータサイエンス時代のDB利用

- クロスボーダーを目指すDB利用者の戸惑い
統合検索の困難
膨大な検索結果
- 求められるものは統合検索の代行と結果の要約解釈？
DB利用者にとってのAIはIntelligent Agent
- DBの利用者はヒトとAI？
Find, Access, Interoperate, Reuseの主語は誰？



バイオデータサイエンス時代のDICP

- いつの時代もコンテンツが勝負
- AI利用は最早コモディティ化
- その先を行くような提案を期待

我が国のバイオサイエンスの状況

- 研究環境の格差が拡大し、研究の機会均等性の喪失
先端研究手法・先端機器へのアクセシビリティには大きな格差
- 研究(者)の多様性の喪失
現行の科学政策に適応した研究者のみが繁栄、それ以外は絶滅の危機？
- 研究(者)多様性保全策としての公共データの統合的利活用環境
研究環境に恵まれない研究者にとっての生存戦略のひとつ
👉 公共性に富む本プログラムのもうひとつの意義

DBを評価する風土・文化の醸成

- データベースの基本は「小確幸」 (©村上春樹)
- 無駄な研究活動の回避という見えない貢献
- データベースの構築と維持はエッセンシャルワーク
- 出口しか評価しない風土・文化からの脱却 (飲水思源)

 評価軸も複線化が必要

バイオデータリソースのサステナビリティ 世界共通の課題



WHAT WE DO ▾ MEMBERSHIP ▾ NEWS & RESOURCES ▾ CONTACT X in

SIGN THE GBC OPEN LETTER

GLOBAL BIODATA COALITION

Biodata resources and the critical ecosystem they form are at risk. As these data continue to grow, we must ensure they are sustained and remain open into the future through effective, long-term financial support.

The GBC convenes the world's research funders to exchange knowledge and share strategies for supporting biodata resources, and provides a discussion forum for the managers of data resources, with the aim of developing principles and models for the coordinated funding of global core biodata resources.

国際協調なくしてバイオデータリソースの維持発展は不可能

世界中のライフサイエンスおよび生物医学研究にとって
長期的な持続性が重要と考えられる52リソースを認定

日本が関与するものとしては、DDBJ、PDB、PXCが選定

将来これらに続くような新しいDBの芽を育成型で支援したい

運営方針

- 研究アドバイザーに加えて、適宜、外部有識者の助言・協力も仰ぎます。
- サイトビジットに加えて、オンライン個別面談も活用して、十分な意思疎通を図ります。
- 本格型も含む全課題参加の対面交流会を開催して、課題間連携の促進による相乗効果の創出を図ります。

最後に

- 生命科学の研究動向を見据えて、広範なユーザーの知識発見・課題解決に資するようなデータベースを、高い志と情熱をもって構築してください。
- 優れた構想、情熱と公共心、そして着実な実行力を兼ね備えた若々しい提案を期待しています。
- 皆さまの積極的な応募をお待ちしております。

提案・選考に関する注意事項など

プログラムのマネージメント

研究総括・アドバイザー



マネジメント



研究開発課題



連携・協業



研究コミュニティ



選考・採択

研究開始

2年次

3年次



公募

構想
立案

節目目標
設定

事後評価

育成型

成果像の明示

- ✓ 利活用・価値の視点で明確化

目標の細分化・密なフィードバック

- ✓ 研究の節目ごとの目標を設定
- ✓ 節目目標は国内外コミュニティのニーズ・利用状況を踏まえ設定
- ✓ 研究開始前にマネジメント側との節目目標を合意

年次レビュー、随時の検証も実施

- ✓ 進捗遅延・目標に向けた道筋からの逸脱を早期把握・介入
- ✓ 有効性検証・利活用例の創出

研究総括・アドバイザーボード

【研究総括】

伊藤 隆司 九州大学 大学院 医学研究院 教授

【研究アドバイザー】（50音順）

鎌田 真由美 北里大学 未来工学部 教授

坂井 寛章 農業・食品産業技術総合研究機構 高度分析研究センター ユニット長

清水 佳奈 早稲田大学 理工学術院 教授

瀬々 潤 (株)ヒューマノーム研究所 代表取締役社長

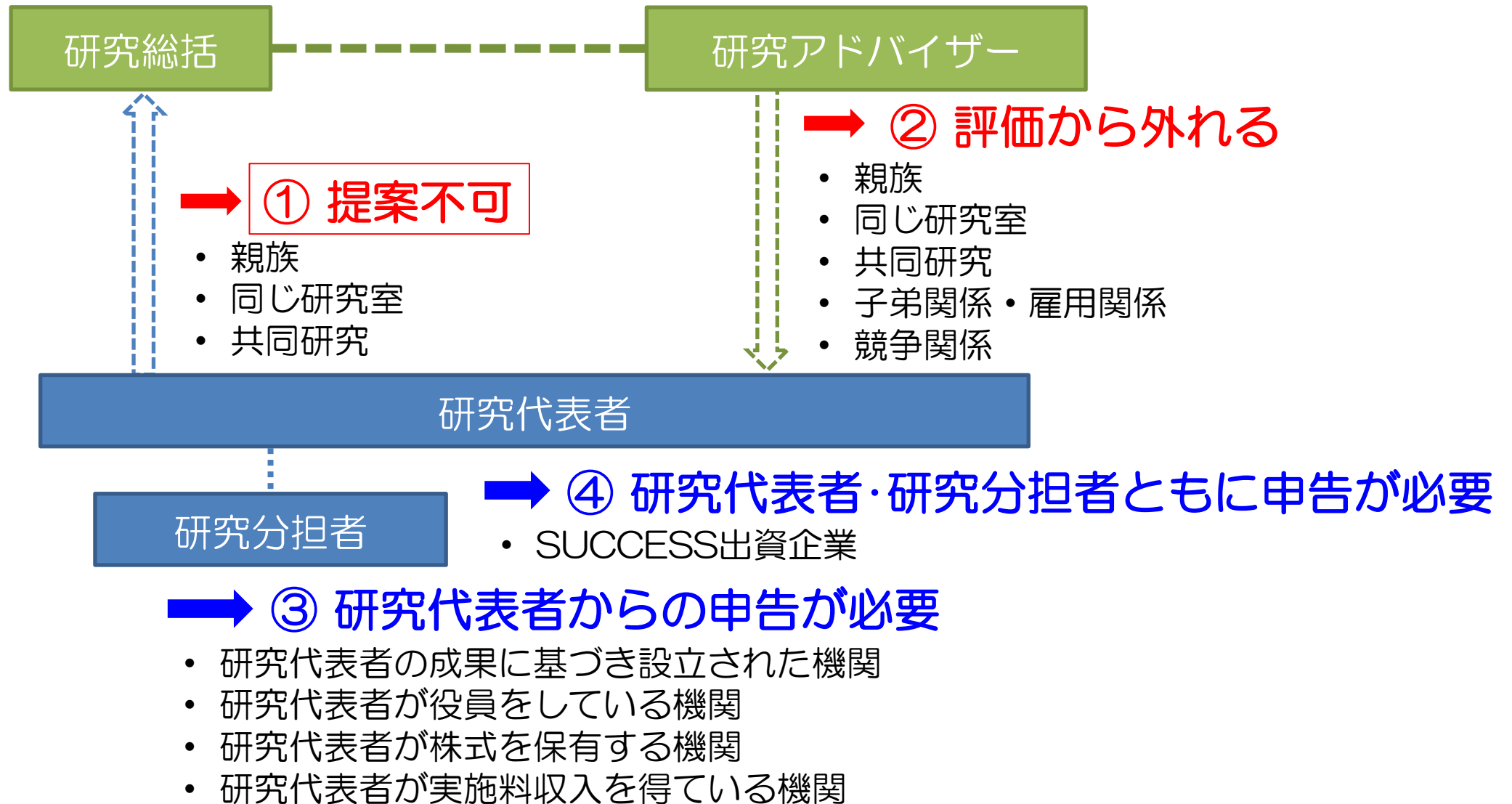
馬場 健史 九州大学 生体防御医学研究所 教授

山本 一夫 お茶の水女子大学 ヒューマンライフサイエンス研究所 客員教授

吉田 哲郎 アクセリード(株) 経営企画部 シニアディレクター

※ 必要に応じ、上記以外の外部有識者に加わって頂く場合があります。

利益相反について



研究開発期間・研究費・採択予定課題数

・研究開発期間

・育成型

2025年4月から3年以内（予定）

・研究開発費（直接経費）

・育成型

～1,000万円／年度程度（予定）

※ 委託研究契約に基づき、研究機関に対して直接経費とは別に、原則として直接経費の30%の間接経費を支払います。

・採択予定課題数

・育成型

～3課題程度（予定）

※ 今回は「本格型」の募集は行いません。

公募期間・選考スケジュール

1. 公募締切： 2025年 1月27日(月) 正午 厳守

【注意事項】

締切り間際はe-Radへのアクセスが集中し、入力できない等のトラブルが発生するおそれがありますので、e-Radでの手続きはなるべく早めに完了するようにしてください。

2. 書類選考結果の通知： 2025年 2月下旬（予定※）

3. 面接選考会： 2025年 3月9日(日)
(都内での対面開催を予定)

4. 選定課題の通知・発表： 2025年 3月下旬以降（予定※）

5. 研究開発の開始： 2025年 4月（予定※）

※ 課題選考手続きの関係で予定よりも遅れることがあります。

ご質問・お問い合わせなど

- ご質問・お問い合わせなどは、メールにてご質問をお送りください。

質問の送り先：nbdc-funding@jst.go.jp

募集説明会の動画とQ&Aの公開

- 質疑応答のパートを除いた募集説明会の動画と説明資料、ならびに本日の質疑応答でご紹介した質問と回答は、近日中に公開いたします。

下記のNBDC公募サイトをご覧ください。

<https://biosciencedbc.jp/funding/calls/2025.html>

皆様からの多数の応募をお待ちしております

問い合わせ先

- JST NBDC公募担当：

- ・本プログラムの内容、応募手続き・利害関係等についての問い合わせ
- ・応募後に他の競争的研究費等に採択された場合の連絡

E-mail：nbdc-funding@jst.go.jp

Tel：03-5214-8491（受付時間：10:00～12:00/13:00～17:00）

※緊急時以外は、メールにて問い合わせ願います。

※土、日、国民の祝日及び年末年始（12月29日～1月3日）を除く

- JST 法務・コンプライアンス部 研究公正課：

- ・不正経理、研究不正、研究倫理教育に関するプログラム等

E-mail：rcr-kousyu@jst.go.jp

- 内閣府 e-Radヘルプデスク：

- ・e-Radの操作方法

<https://www.e-rad.go.jp/contact.html>

Tel：0570-057-060（受付時間：9:00～18:00）

※土、日、国民の祝日及び年末年始(12月29日～1月3日)を除く