LINC のワーキング紹介

現在 10 のワーキング（WG）があり、各 WG の中ではタスクフォース（TF）とプロジェクト（PJ）が活動しています。

1 WG00 : AI/DB 基盤
2 WG01 : デジタルヘルス&メディスン、RWD 解析
3 WG02 : 標的探索・バイオインフォマティックス
4 WG03 : バイオロジックス・ニューモダリティー
5 WG04 : 分子シミュレーション
6 WG05 : メディシナルケミストリー
7 WG06 : 製剤・マテリアルズインフォマティクス
8 WG07 : ADMET・トランスレーショナルリサーチ
9 WG08 : 治験・市販後・メディカルアフェアーズ
10 WG09 : 人材育成

LINC の WG（ワーキング）

WG02
標的探索，バイオインフォマティクス
- 標的探索（実測データ、文献、DB）
- 薬剤→標的分子→パスウェイ→疾患

WG04
分子シミュレーション
- 構造生物学 AI
- ドッキング AI
- 分子動力学 AI
- AI 力場

WG08
治験・市販後
- 治験の効率化
- 費用対効果分析
- 治療最適化 AI モデル
- 複数 RWD の統合 DB

WG01
デジタルヘルス&メディスン、RWD 解析
- マイクロバイオーム
- 発症予測
- デジタルサロゲートマーカー

WG03
バイオロジックス・ニューモダリティー
- 抗体物性
- アプタマーデザイン
- アンチセンス核酸
- ニューモダリティ

WG05
メディシナルケミストリー
- 農薬　DB、予測モデル
- 化合物デザイン
- モデル精度向上
- 合成経路

WG06
マテリアルインフォマティクス
- 製剤処方設計 AI
- データベース構築
- 溶出シミュレーション
- ラボノート & エッジ AI

WG07
トランスレーショナルリサーチ
- 解析自動化システム
- モデル動物提案システム
- 臨床薬効予測基盤開発
- 経路予測の高度化

WG09　人材育成

WG00　AI / DB 基盤

WG00

AI/DB 基盤

ライフサイエンス分野における AI 開発とビッグデータの活用を促進するために、基盤となるデータ整備、アルゴリズム開発、ハードウェア整備に加え、医療・医薬品データの提供と利用推進にむけた提案などを行います。

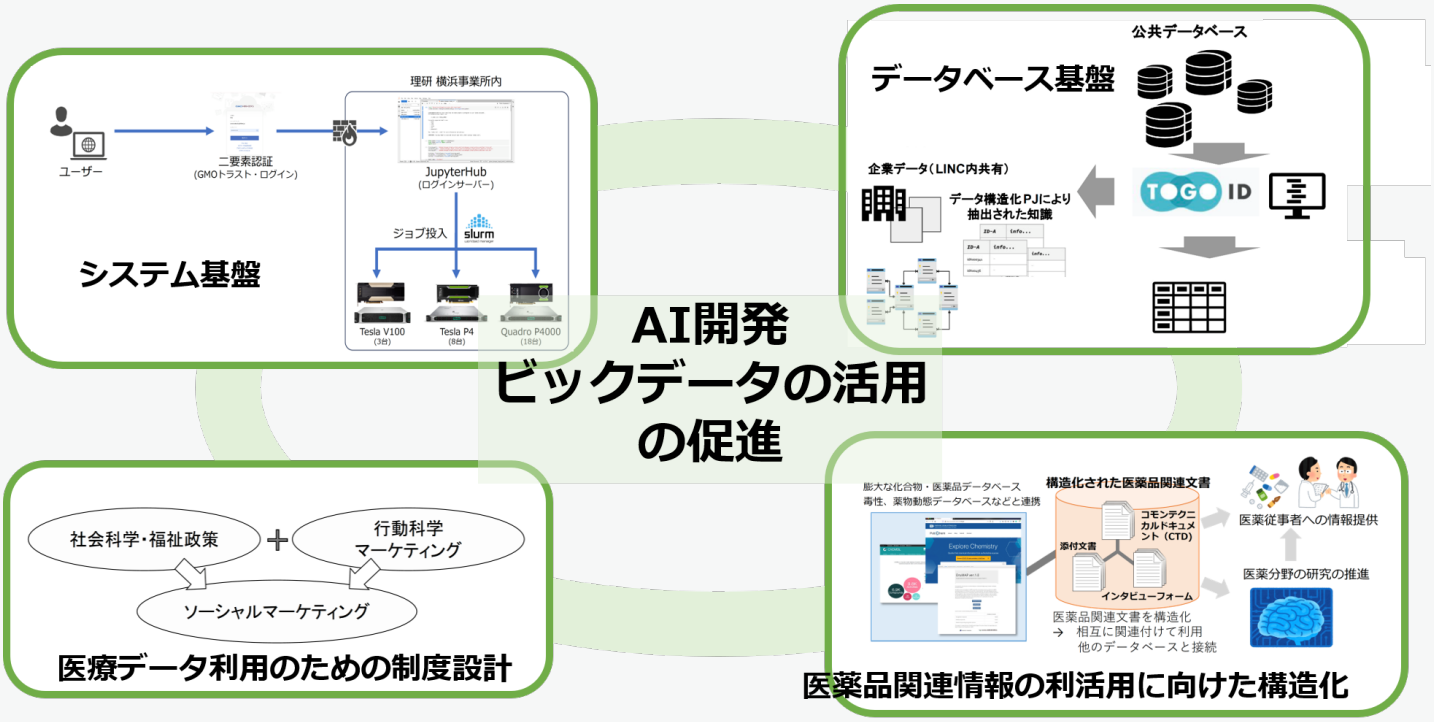
【 WG リーダー 】
池田 和由　理化学研究所
長尾 知生子　大阪大学蛋白質研究所／(国研）医薬基盤・健康・栄養研究所

■ タスクフォースとプロジェクト

1	WG00-tf-fedrated	AMED DAIIA 関連（日本版 MELLODDY, 秘密計算 , システム基盤）
2	WG00-tf-meddata	医療データ収集、提供のための調査・分析と制度設計、基盤構築
3	WG00-tf08	医薬品関連文書の有効活用に向けたフォーマットの検討と提案（WG08-tf05 と共同）
4	WG00-tf09	統合データ共有基盤の構築

AI 開発とビッグデータの活用の基盤となるデータ整備、アルゴリズム開発、ハードウェア整備に加え、医療・医薬品データの提供と利用推進にむけた提案などを行っています。
理化学研究所内の計算機クラスターに、ウェブブラウザから AI開発が可能なセキュアな開発環境を構築し、活用いただくことで、LINC 内での AI開発のサポートと成果利用・評価サイクルの効率化を目指しています。また、医療データや医薬品関連情報の利活用に向けた活動も行っています。医療情報提供の制度設計の提言にむけたアンケートを実施したほか、インタビューフォームを構造化する試みも行っています。

ライフサイエンス統合データベースセンターと共同で、公共データベースおよび、LINC 内で構築される知識データを紐づけ、シームレスに活用可能なデータ基盤の構築に向けた活動もしており、必要に応じて各 WG の PJ との連携を進める予定です。他に自然言語処理や量子コンピュータなど、WG 横断的な技術基盤の整備が必要となった場合には、臨機応変に対応していきます。



WG01 デジタルヘルス&メディスン･RWD 解析

WG01 では、未病・疾病予測・予防の社会的ニーズに応えるため、腸内細菌、健診、食事、歩行などのリアルワールドデータを収集、活用する AI プラットフォームを作成し、モバイルを活用した社会実装を目指します。

【WG リーダー】
内野 詠一郎／京都大学大学院　医学研究科
荒木 通啓／（国研）医薬基盤・健康・栄養研究所
伊藤 眞里／大阪大学大学院　薬学研究科

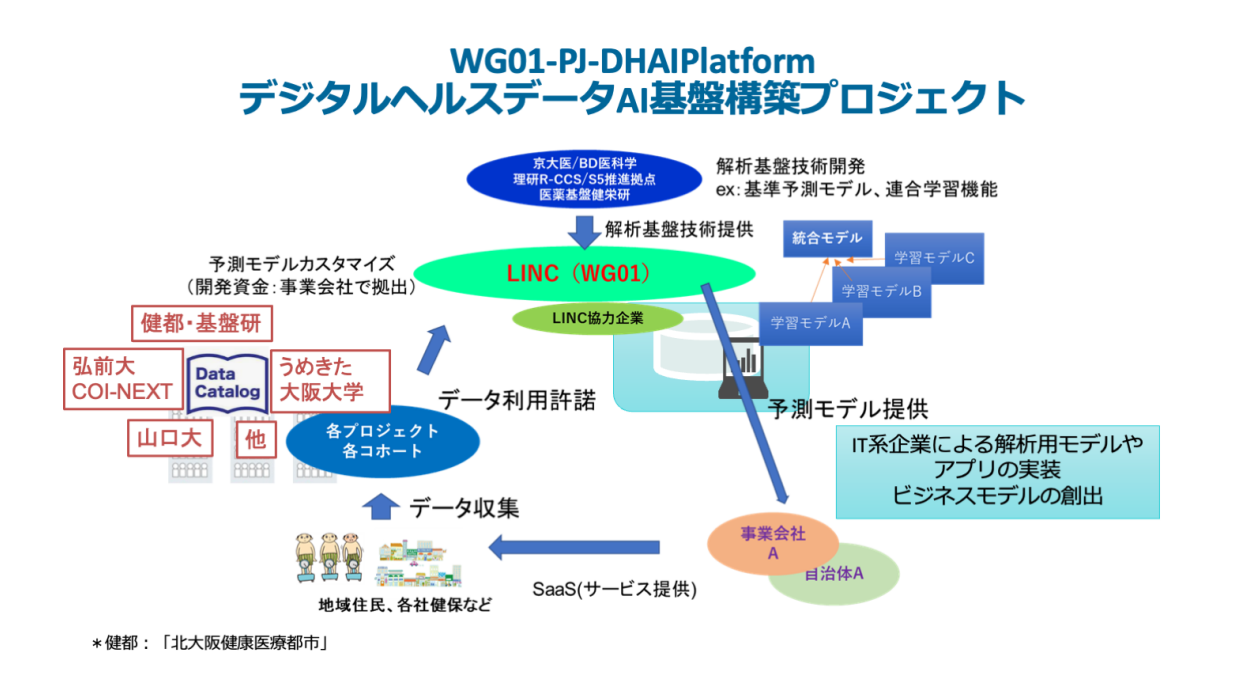
■ タスクフォースとプロジェクト

1 WG01-tf01　健康情報&マイクロバイオーム・オミクスデータによる未病・先制医療への貢献
2 WG01-tf02　健康診断データによる発症予測モデルの社会実装
3 WG01-tf03　SNS データによる発症因子同定とリスク予測
4 WG01-tf04　食事カロリー計算の AI モデルの構築
5 WG01-tf05　メンタルヘルスの予測アプリ
6 WG01-tf06　歩行解析による運動器疾患リスク予測アプリ
7 WG01-tf07　SNS データによる発症因子同定とリスク予測
8 WG01-PJ-DHAIPlatform　デジタルヘルスデータ基盤構築

WG01 では、上に示すような様々な AI 実装イメージに沿った健康医療データセットの選定やモデル開発の方向性について議論してきました。いずれの議論においても、現状では健康医療データの取得内容、拠点やデータ形式が散在していることなど、利活用の枠組みが未だ充分に整っていないことが課題として挙がっています。

WG の中心的な活動として、現在はデジタルヘルスデータ基盤構築プロジェクト（WG01-PJ-DHAIPlatform）において様々なデータセットのカタログ化と、それらをもとにした統合的に扱える AI 基盤の構築を目指した活動をしています。図に示すような日本全国各地にわたる健診データの取得拠点先とも連携しデータの統合利用のための情報整理と国際的な動向も踏まえた調査検討を進めています。

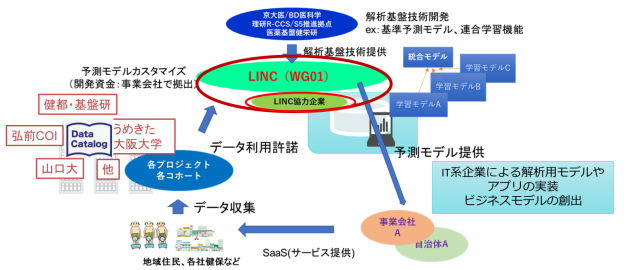
また 2023 年度後半からは、JST 共創の場形成支援プログラム（COI-NEXT）の弘前大学拠点での連携拠点として本 PJ が採択され、公的な立場としても活動を行っています。アカデミア、ドメイン企業、IT 企業の様々な視点から、健康医療データによるヘルスケア DX を推進していきます。



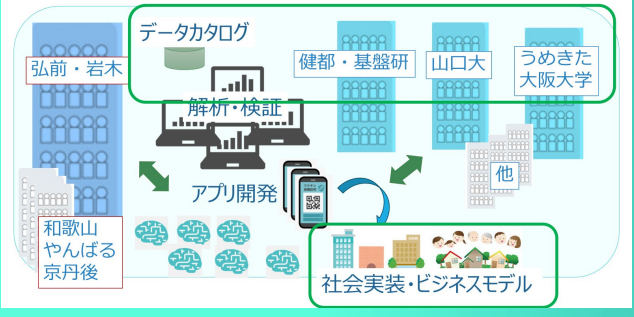
WG01-PJ-DHAIPlatform(デジタルヘルスデータ基盤構築)
全国各地のコホートデータとの連携プラットフォームを目指して

【PJ設置の背景と概要】
弘前COI、健都、山口大学をはじめ、全国各地のコホート研究においてデジタルヘルス(DH)データ収集が盛んであり、これら現行のフィールドのニーズに添いながら、LINCの利点を活かし、データ収集→統合→AI解析→解釈までを一気通貫で行えるプラットフォームを構築する。
このプラットフォームは、将来的にヘルスケアDX基盤構築プログラムとして展開を図る。

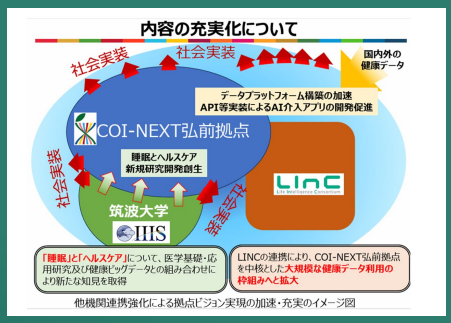
デジタルヘルスデータ基盤構築プロジェクト



弘前COI-NEXTへの連携によりプロジェクトが加速



進捗および今後の予定
本PJが弘前COI-NEXTへの連携機関として採択された



プロジェクトのタスク＆スケジュール

	2023	2024	
データ基盤構築	進行中	計画	ゴール
	・連携先各コホートの項目一覧、詳細、データ形式情報の収集	・連携先各コホートとのデータ統合利用の調整・環境整備	・コホートカタログ実装
	・カタログフォーマットへのキュレーションや階層構造の整理 ・主要拠点データ項目との統合性検討	・生成AIを活用したデータ統合作業	・統合データ取得・利用システムの整備完了
	・統合利用のためのデータ構造、メタデータ仕様の決定		
	・各コホートのデータへのアクセス方法、セキュアな提供手法など、諸外国のシステムも参考にして調査	・民間企業等での利用スキームの検討	・AIモデル＆アプリ搭載プラットフォーム完成
アプリ開発と実装			
	・AIプログラムのチューニングとアプリ開発	・参画企業・機関での実装研究の企画と推進 ・AIアプリの実装フィールドにおける検証	

今後のアクションまとめ

①コホートデータ項目を階層化した
カタログ作成と実装
②タスクフォースの連携、再稼働へ準備
③ヘルスケアAI基盤モデルの開発構想の具体化
④AIプラットフォームを中心にした
デジタルツイン体制の構築

大規模健診データ学習によるヘルスケアAI基盤モデルの開発

大規模健診データ学習によるヘルスケアAI基盤モデルへ
（言語だけでは個人の健康状態の予測や改善は不可能）

WG01 タスクフォース一覧

tfNo.	タイトル
tf01	健康情報＆マイクロバイオーム・オミクスデータによる未病・先制医療への貢献
tf02	健康診断データによる発症予測モデルの社会実装
tf03	臨床応用可能なデジタルサロゲートマーカーの検討
tf04	食事カロリー計算のAIモデルの構築
tf05	メンタルヘルスの予測アプリ
tf06	歩行解析による運動器疾患リスク予測アプリ
tf07	SNSデータによる発症因子同定とリスク予測

PJメンバー

リーダー	
京都大学	内野詠一郎
医薬健栄研	荒木　通啓
大阪大学	伊藤　真里

ご所属	お名前	ご所属	お名前
山口大学	中津井　雅彦	Data Robot Inc	伊地知　晋平
下関市立大学/山口大学	中上　裕有樹	中外製薬株式会社	伊川　由利子
大阪大学	長尾　知生子		(岡田　法大)
医薬健栄研	瀧本　秀美	大塚製薬株式会社	金子　和裕
	松本　麻衣		古川　恵実
	岡田　知佳	EAファーマ（株）	大場　史記
	深川　明子		黒須　真介
	樋口　千洋	第一三共株式会社	佐藤　和明
理研	安倍　理加	サラヤ株式会社	阪長敏朝郎
弘前大学	玉田　嘉紀	株式会社エクサウイザーズ	安田　英史
	橋本　泰裕	株式会社ちゃんフク	福田　宏幸
	岩根　拓朗		

WG02

標的探索・バイオインフォマティックス

WG02 では、幅広いモダリティーでオミクスデータや既知情報を活用し、データ駆動的に疾患の発症メカニズム推定や創薬標的候補の提示を可能とする技術の開発を行います。

【 WG リーダー 】
夏目 やよい／（国研）医薬基盤・健康・栄養研究所
玉田 嘉紀／弘前大学

■ タスクフォースとプロジェクト

1 WG02-PJ-omics　実測データ（オミクスデータ）からの標的探索
2 WG02-PJ-knowledge　ナレッジグラフからの薬剤併用効果予測

WG02 では、データ駆動型で創薬標的探索や薬剤作用機序および疾患発症機序解明を、AI技術を用いて行っています。現在は2つのプロジェクトが進行中で1つはオミクスデータからの標的探索、もう一つはナレッジグラフからの薬剤併用効果予測です。

オミクスデータからの標的探索では、ベイジアンネットワーク推定技術を用いてオミクスデータ、主に遺伝子発現データを用いて遺伝子ネットワークを推定・解析することで新規薬剤標的の探索を目指しています。ベイジアンネットワークは、因果関係の仮説として、変数間の関係性を人間が理解できるネットワークという形で示すことにできる説明可能 AI の一つです。

大規模なデータセットからこの技術を用いて遺伝子ネットワークを推・解析するにはスーパーコンピュータを用いた計算が必要なことからハードルが高い技術ですが、本 WG では理研のスパコン富岳や東大医科研のスパコン SHIROKANE を利活用する技術を持っているメンバーが揃っており、現在は肝線維化という具体的なテーマをターゲットに公共データを用いて実証を行っています。

スーパーコンピュータを用いた解析技術は単独企業では手を出しにくい技術でもあり、LINCとしてこれに取り組むことで、新しい解析法の理解や普及、並びに参加メンバー各社の技術力の向上も図ることができ、LINCならではの取り組みができていると自負しております。

ナレッジグラフからの薬剤併用効果予測では、様々な医学的データ、生物学的データからナレッジグラフと呼ばれる構造（情報を点と線で表現することで、異なる情報がどのように関連しているかを AIで解析しやすい形に変換）を作り、グラフに対して適用できる AIの学習データとして用いることにより、これまで知られていなかった関係性を予測することを目指しています。

例えば、1剤で引き起こしうる副作用については比較的多くの情報がありますが、複数の薬を服用する場合に引き起こしうる副作用については情報が限られています。

そのため、1剤で引き起こしうる副作用の情報に関連する医学的データ、生物学的データを加えてナレッジグラフを構築し、ここから「まだ知られていない、薬剤を併用することで引き起こしうる副作用」を予測する AIを構築しています。

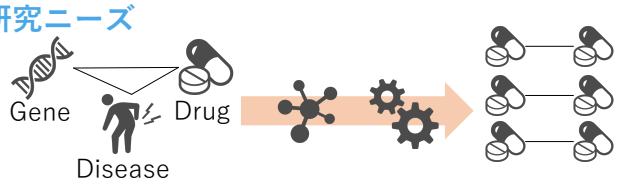
これらの成果については学会発表を行うなど、アクティブに研究活動を進めています。製薬企業、IT企業、アカデミアがそれぞれ異なる技術や知識を持ち寄り、組織の壁を越えた有機的な連携をすることにより、AI技術の社会実装における課題や創薬における問題意識を皆で共有しながら切磋琢磨する場となっています。

WG02-pj-knowledge

Drug Response and Synergy Prediction using Knowledge Graph Technology

Introduction｜入力グラフで結果が激変!?

研究ニーズ

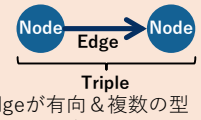


併用することで副作用が発現する・併用するとシナジーが生じる薬の組合せを予測したい

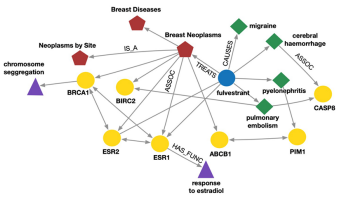
ナレッジグラフで解決したい！

生体・疾患・薬はネットワーク的な関係であり、問題解決にはグラフ技術が適している。

グラフ：
データ間の関係を表現



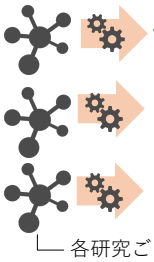
Edgeが有向＆複数の型
⇒ナレッジグラフ



Example: Bio Knowledge Graphs

技術的な課題

しかし…データセットが少しでも異なると、グラフでは予測結果に大きな違いが生じる(かも)。



予測結果
予測結果
予測結果

単純に比較できない
- 本当にいいアルゴリズムが分からない

考慮すべき要素は多い
- ラベルの誤り
- インバランス
- 欠損値
- Negativeデータ など

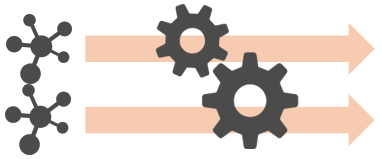
└ 各研究ごとに独自に作成されたデータセット

目的とゴール

グラフMLのための“よりよい”データセットかを測る指標を提案

MLに適したデータセットを構築するための指標＆異なるデータセットでの比較の一助とする

Method｜複数データセット×同じアルゴリズムで予測し、違いを検討する



予測結果
予測結果

データセットの違い　　予測結果の違い

データセットの違いがどのように予測結果に影響を与えているかを考察。何らかのデータセット評価指標を作れないか検討する。

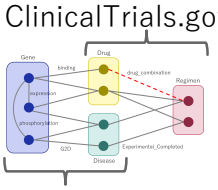
So far｜昨年度

- 併用による副作用予測を実施
  Decagon[2]、TargetMine[3]＋Polypharmacy Side Effect予測データ[4]など複数データセットで試行
- データセットに副作用以外のデータも多数含まれており、副作用以外の予測結果も混入

So far｜2024：薬のシナジーにフォーカス

データソース: ClinicalTrials.gov[5]＋TargetMine

ClinicalTrials.gov

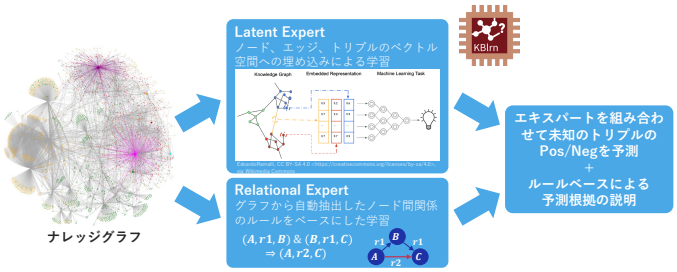


TargetMine

終了した全疾患領域の薬物介入試験データを使用
- 表記ゆれに対応した上でTMへ名寄せ
- Active Comparatorおよび中止せず終了した試験のExperimental群レジメン
  (複数ある場合は最新の値を採用)

以下データおよびそれぞれを繋ぐ相互作用データを使用
- 低分子化合物 – 遺伝子 – 疾病
- 化合物 – 遺伝子
- 遺伝子 – 疾病

グラフリンク予測アルゴリズム：KBlrn



予測精度評価（情報科学的な評価）

	MR	MRR	Hits@1	Hits@3	Hits@5	Hits@10
CT.gov	166	0.228	0.119	0.245	0.345	0.482
CT.gov+ TargetMine	256	0.200	0.105	0.232	0.299	0.391

CT.govでGroundTruth(GT)が上位で予測されたレジメンでも、化合物・遺伝子・疾病間の相互作用データを学習することにより、CT.gov+TargetMineでは異なる薬が上位で予測された

Discussion｜結果評価とこれから

1. **精度指標については十分な結果だった**
2. **定性評価で見られた傾向：**①インバランス、②GTと同じATC分類の薬が上位で予測される、③TMによりVEGF抗体薬/ICI/mTOR阻害薬が上位で予測された
   - **インバランスの問題：**臨床試験でよく使われる薬（化学療法薬）の影響が強く出ている
3. **GrondTruthと異なるATC分類の薬が出力された場合の定性評価（予測によって構成されたレジメンの臨床的意義）：**TMにより適応症に対し有効性を示しうるレジメンも予測された例: [‘brivanib’,‘fluorouracil’,‘cisplatin‘] (GT= 'leucovorin’)
4. **大目的に向けて：**生物学的に正しいと分かっているスコアに対して近いものは何かを探索。機械学習でのもっともらしさと、生物学的なもっともらしさを比較する予定
   - 生物学的なもっともらしさ＝ATC分類と仮定し、GT×予測結果の分布を比較

1) Jure Leskovec, Stanford CS224W: Machine Learning with Graphs, http://cs224w.stanford.edu, 2)Marinka Zitnik et al., Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics*, Volume 34, Issue 13, 01 July 2018, Pages i457–i466. 3) Chen YA et al., TargetMine 2022: a new vision into drug target analysis., *Bioinformatics*. 2022 Sep 15;38(18):4454-4456. 4)Zheng et al., Predicting adverse drug reactions of combined medication from heterogeneous pharmacologic databases. *BMC Bioinformatics* 19 (Suppl 19), 517 (2018).

LINC SHOWCASE 2024/3/21

WG03

バイオロジックス・ニューモダリティー

WG03 では、低分子医薬品では制御することが難しい標的分子に対して、核酸医薬品・抗体医薬品・その他のニューモダリティを利用した医薬品開発に関連するインシリコ創薬技術（AI、シミュレーション等）確立を目指します。

【 WG リーダー 】
千葉 峻太朗／理化学研究所

■ タスクフォースとプロジェクト

1 WG03-tf-ab　抗体 (antibody) の親和性改善 or 維持・物性改善についての課題と取組み方について検討
2 WG03-tf-am　アプタマーデザインの課題と取組み方について検討
3 WG03-PJ-as　Gapmer, siRNA, スプライシング制御型アンチセンス核酸の課題と取組み方について検討

■ バイオロジクス・新規モダリティを利用し、医薬プロセスを最適化を目指しています

バイオロジクス・新規モダリティによって、低分子で標的にすることが難しい分子の機能を制御するために有用であると期待

	【ペプチド医薬品】	【核酸医薬品】	【タンパク質医薬品・抗体医薬品】
分子（例）	・天然分子模倣のペプチド ・新規ペプチド（サイクリックペプチドなど） ・コンジュゲートペプチド	・アンチセンス核酸（スプライシング制御、Gapmer） ・siRNA ・アプタマー	・ネイキッド抗体 ・Bi-specific 抗体 ・ADC ・小型化抗体 ・De novo タンパク質
課題（例）	・ワクチン ・膜透過化 ・アミノ酸構造改変 ・徐放製剤 ・代謝安定化	・配列デザイン ・人口核酸の選択 ・DDS **アンチセンスプロジェクト** **アプタマータスクフォース**	安定性・親和性・物性の改善を目指したアミノ酸改変、バッファーの選択 **抗体タスクフォース**

WG04

分子シミュレーション

WG04 では、創薬標的タンパク質や、医薬品候補の化合物の立体構造に基づく計算創薬（分子シミュレーション、人工知能応用）に取り組みます。スーパーコンピュータ富岳を活用する TF もあります。

【 WG リーダー 】
池口 満徳／横浜市立大学 生命医科学研究科

■ タスクフォースとプロジェクト

1 WG04-PJ-aiff　機械学習を用いた高精度分子力場の構築
2 WG04-PJ-ihbvs　相互作用と水和情報に基づく仮想スクリーニング
3 WG04-PJ-KBDD　スパコンを用いたドラッグディスカバリー
4 WG04-PJ-MDAI　分子動力学 AI
5 WG04-PJ-stbiol　構造生物学 AI
6 WG04-PJ-uldoc　ドッキング AI、富岳 S5

■ WG04 分子シミュレーションでは、分子レベルでの計算創薬に関わる研究開発を実施しています

PJ-AIFF では、量子化学計算結果から高精度分子力場を学習する AI力場を開発しています。PJ-IHBVS では、理研独自の相互作用記述子と標的タンパク質の水和情報に基づく新たな仮想スクリーニング技術の開発を進めています。

PJ-KBDD では、スパコン「富岳」を用いた計算創薬研究を行っています。PJ-MDAI では、分子動力学シミュレーションの結果から、自動的に重要な相互作用等を見出す AIの開発を進めています。

PJ-STBIOL では、X線やクライオ電顕の低解像度の実験データを元に、立体構造の品質評価および構造修正を行うことができる AIの開発に取り組んでいます。

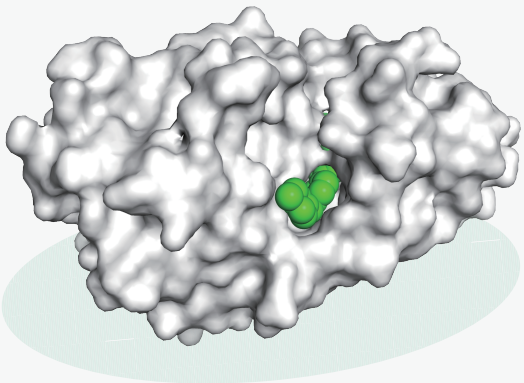
PJ-ULDOC では、Society 5.0 プロジェクトと連携し、スパコン「富岳」を用い、ウルトララージサイズの化合物ライブラリに対するドッキングプラットフォームの開発を行っています。

■ 分子レベルの AI、シミュレーション研究を実施

IHBVS
相互作用・水和情報による
バーチャルスクリーニング

AIFF
AI による高精度分子力場

ULDOC
ウルトララージライブラリ
バーチャルスクリーニング



KBDD
スパコン「富岳」での計算創薬

STBIOL
構造生物学 AI

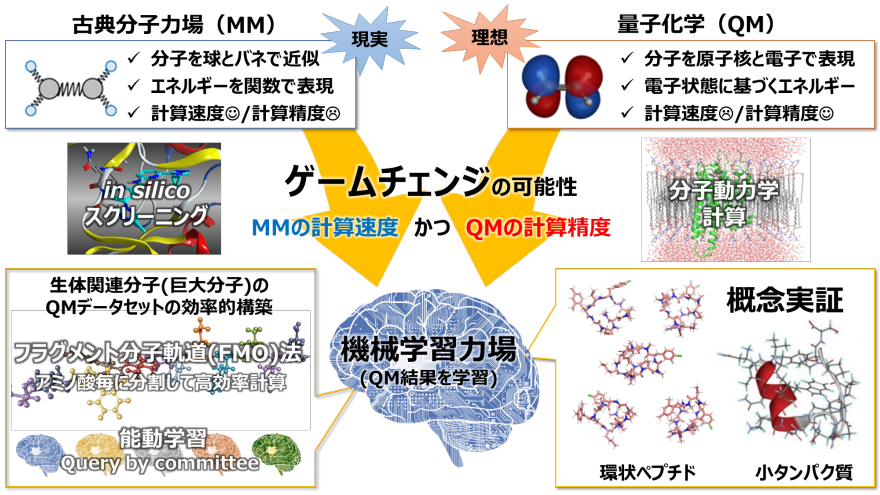
MDAI
分子動力学 AI

WG04-pj-aiff
機械学習を用いた高精度分子力場の構築

PJメンバー/概要

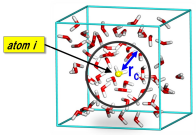
氏名	所属
加藤　幸一郎（PL）	九州大学
増田　友秀（サブPL）	東レ
大島　勘二	カネカ
小澤　基裕	キッセイ
原田　俊孝	住友化学
谷村　直樹	みずほR&T
奥脇　弘次	JSOL

実動部隊：
九大学生(M1:D進予定, B4)をアサイン！！
アドバイザー：
池口先生、大田先生、本間先生、渡邉千鶴先生



機械学習モデル

- N入力(ACSFs)1出力(原子エネルギー)のNNの組み合わせ
- NNは基本的には、原子種ごと(酸素用、水素用、・・・)に用意
- NNP:原子エネルギーを足し上げた全エネルギー(ENN)とEQMのMSE最小化
  NNおよびACSFsは解析微分可能→学習済ネットワークの微分から力を算出



着目原子の周辺環境を並進・回転対称なベクトルに変換(ACSFs)

Radial symmetry functions

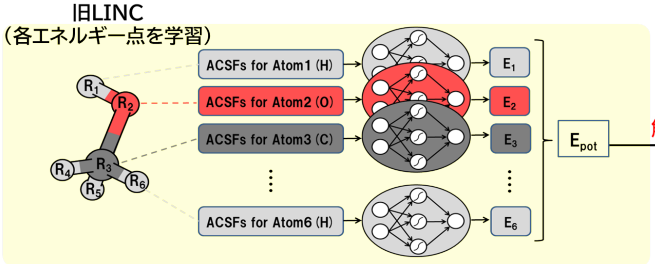
$$G_i^2 = \sum_{j \neq i} e^{-\eta(r_{ij}-r_s)^2} f_c(r_{ij})$$

Angular symmetry functions

$$G_i^4 = 2^{1-\zeta} \sum_{j \neq i} \sum_{k \neq i,j} \left[(1+\lambda\cos(\alpha_{ijk}))\right]^{\zeta}$$

cutoff functions

$$f_c(r_{ij}) = \begin{cases} tanh^3\left[1-\frac{r_{ij}}{r_c}\right] & with\ r_{ij} \leq r_c \\ 0 & with\ r_{ij} > r_c \end{cases}$$

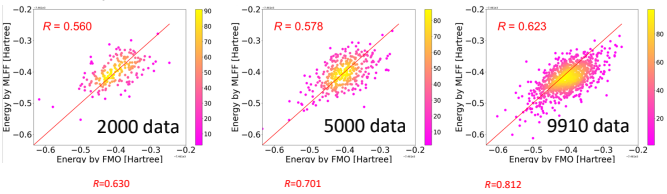


進捗状況

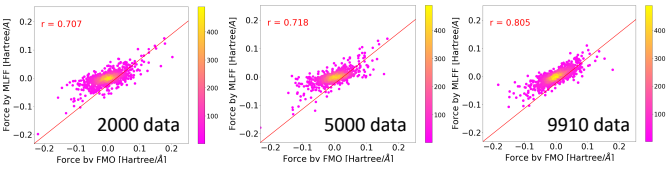
機械学習力場(FMO-AI力場)

◆Trp-Cageでデータ追加(FMO-MP2/6-31G*)

■ エネルギー



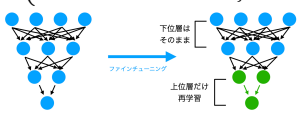
■ 力

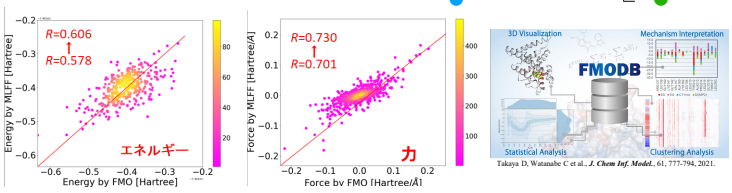


◆FMODBを用いた転移学習

■ 15434エントリーのデータで事前学習

■ Trp-Cage 5000データで転移学習

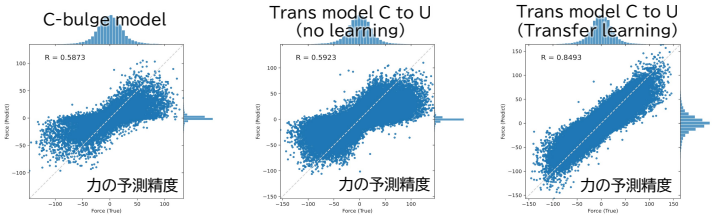




◆RNAの試行

■ 25-mer RNA(バルジあり)を対象に学習

■ C-バルジとU-バルジの双方を比較(両者の転移学習も実施)



今後の展望

- ◻ 転移学習 ⇒ FMODBデータ増で事前学習 ⇒ 力を含めたTrpCageデータなどで転移学習
- ◻ 能動学習 ⇒ ハイパーパラメータ違いの複数の予測モデル構築 ⇒ 予測結果のバラつきが大きい未知構造をFMO計算してデータ追加
- ◻ ハイブリッドAI力場 ⇒ 古典力場とFMOの差分を学習して、古典力場に加算

参加者は随時募集中です！！

謝辞

本プロジェクトはFMO創薬コンソーシアム(https://fmodd.jp/)との連携の下で推進中です。
計算リソースには富岳((project IDs: hp230131)及びHOKUSAI(ID: Q22306)を使用させて頂きました。

hp230025:新薬開発を加速するインシリコ創薬基盤の構築 –タンパク質–ペプチド医薬品/核酸–低分子医薬品の結合ポーズを推定する方法論の構築–

課題ID hp120048, hp140042, hp150025, hp160010,hp170036,hp180011,hp190020, hp190021, hp200011,hp210048, hp220026 後継プロジェクト

bioGrid　WG04 PJ KBDD 新薬開発を加速するインシリコ創薬基盤の構築(「富岳」産業課題)

京都大学大学院・医学研究科/奥野恭史

KBDD bioGrid

●「京」・「富岳」による製薬業界のオープンイノベーション

KBDDコンソーシアム: "K" supercomputer-based drug discovery consortium

申請主体:研究代表 (一社)ライフインテリジェンスコンソーシアム 奥野恭史
事務局: NPO法人バイオグリッドセンター関西
製薬企業(11社):
科研製薬(株)、(株)カネカ、キッセイ薬品工業(株)、住友化学(株)、千寿製薬(株)、大正製薬(株)、田辺三菱製薬(株)、中外製薬(株)、東レ(株)、日本新薬(株)、マルホ(株)
IT企業(1社):三井情報(株)
大学等: 京都大学医学研究科, 横浜市立大学,大阪大学蛋白質研究所、筑波大学, 兵庫県立大学、千葉大学医学部、獨南大学、山口大学、医薬基盤研究所、(公財)都市活力研究所

- コンピュータ創薬の根本課題に挑戦
- 製薬会社による現場利用に耐えうる計算フロー(計算精度と計算時間)の構築
- 我が国のコンピュータ創薬の中心拠点形成

●「京」・「富岳」による医薬品開発効率の抜本的解決

実験の代わりにコンピューター上で結合をシミュレーションする ⇒ 開発費用の削減への期待

3000万化合物　10万-10000化合物　1000-100化合物　100-10化合物　10-1化合物

CGBVS法の適用範囲　MP-CAFE法の適用範囲

CGBVS法:Chemical Genomics Based Virtual Screening
タンパク質と化合物の組合せを計算するためパターン認識技術を適用
世界最大規模(189.3億ペア)の結合予測
631種の疾患原因タンパク質と3000万種の化合物
予測結果を基に、製薬会社20社が独自に開発に利用中

MD/MM-PBSA法:
Molecular Dynamics / Molecular Mechanics Poisson-Boltzmann Surface Area
水分子も含めて溶質の動きをシミュレートする分子動力学計算と結合自由エネルギー計算の組み合わせによる、リガンド結合ポーズの予測
水溶性タンパク質32種と阻害剤の結合ポーズを予測
水溶性タンパク質32種×阻害剤数15種の組み合わせについて、化合物ドッキング+MD/MM-PBSAおよびマルチカノニカルMDを適用し、化合物結合ポーズの予測精度を検証した。

MP-CAFEE法:
Massively Parallel Computation of Absolute binding Free Energy
アルケミカル置自由エネルギー摂動法により、結合親和性の精密予測を実現
キナーゼ・プロテアーゼ・GPCRと阻害剤の結合親和性を予測
CHK1キナーゼ6化合物、CDK2キナーゼ-16化合物、ERK2キナーゼ-11化合物、Urokinase-7化合物、Adenosine A2a receptor-5化合物について結合自由エネルギー(ΔG)を求めたところ、結合定数の実験値(Kd)との間に良好な相関関係がみられた。

MD/MSM法:
Molecular Dynamics / Markov State Models
複数の短時間MD(分子動力学計算)をMSMに基づいて統合することで分子構造の長時間ダイナミクスを推定し、リガンド結合過程におけるk_{on}, k_{off}を予測
ジヒドロ葉酸レダクターゼ(DHFR)、熱ショックタンパク質(HPS90)、キナーゼ(CDK2)、GPCR(β2AR)の結合経路を予測
仮説状態から始めた多数の分子動力学シミュレーション(100ns)を統合しMSMを構築し、結合自由エネルギー(ΔG), 結合速度定数(k_{on}), 解離速度定数(k_{off})を求めた。また、100ns程度のMDシミュレーションでは結合部位を観測できないタンパク質-リガンド系に対しては、マルチカノニカル法とMD/MSM法の組み合わせを試みた。

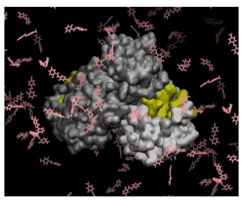
①結合するかどうかの予測　②結合ポーズの予測　③結合の強さの予測　④k_{on},k_{off}の予測

・H24-25年度　実施済み　・H27-28年度　実施済み　・H24-26年度　実施済み　・H29-R2年度実施済み

● R3年度の研究内容 「分子動力学シミュレーションに基づいた次世代インシリコスクリーニング手法の開発」

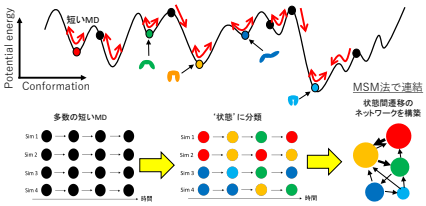
(目標)富岳の高い並列コンピューティング能力を活用する事で、大量の化合物に対して精密な結合シミュレーションを実施し、タンパク質表面上の化合物結合サイトを網羅的に探索すると共に、各結合サイトに対してΔG, kon, koffを精密に推定する新規バーチャルスクリーニング手法(ColDock/MSM法)を開発する。

ColDock法:東京工業大学・北尾彰朗博士によって開発された手法であり、標的タンパク質の周囲に低分子化合物を高濃度に配置した状態から分子動力学計算を実施する。本法においては、化合物同士の間に反発力を導入しているため、化合物同士の凝集を防ぐことができ、タンパク質への結合解離イベントを効率的にサンプリングできる。

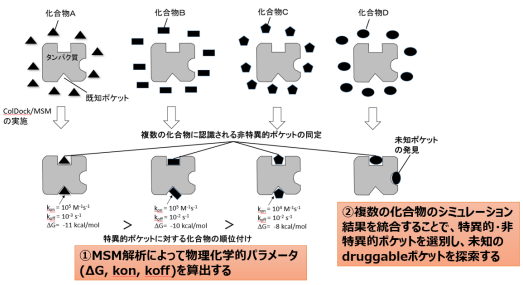


MSM(Markov State Models)法の原理

複数の短時間MD(分子動力学計算)を統合し、構造クラスタリングにより定義した状態間の遷移モデルを構築する事で、分子構造の長時間ダイナミクスを推定する解析手法



<ColDock/MSM法による化合物スクリーニングの概念図>

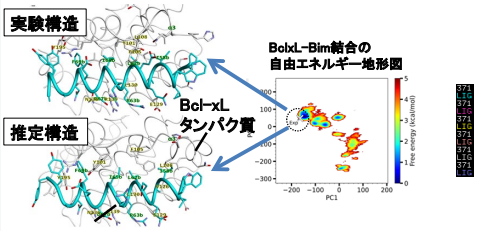


● R4-5年度の研究内容

「タンパク質-ペプチド医薬品/核酸-低分子医薬品の結合ポーズを推定する方法論の構築」

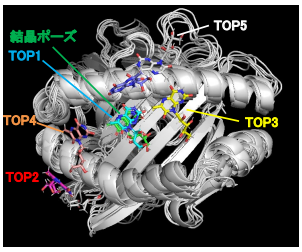
<R4年度の研究成果>

McMD[1]によって推定したBcl-xL-Bimペプチド複合体構造



<R5年度の研究成果>

gREST[2]によって推定した抗原提示タンパク質と医薬品の複合体構造



[1] Nakajima, N.et al. J. Phys. Chem. B 101, 817–824. (1997)
[2] Kamiya M. et al, J. Chem. Phys. 149, 072304 (2018)

● R6年度の研究内容

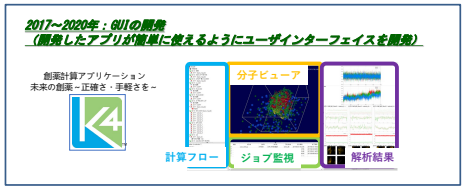
「実践的な創薬テーマを通したインシリコ創薬基盤の評価・精密化」

(概要)近年、製薬業界では医薬品の研究開発費が高騰しており、計算機シミュレーションによる創薬(インシリコ創薬)に大きな期待が寄せられている。我々はこれまでに、McMD法・gREST法といった拡張アンサンブル分子動力学シミュレーションに基づいて、タンパク質に対する低分子医薬品及びペプチド医薬品の結合ポーズを精密に推定する方法論を開発・実装し、代表的なベンチマークタンパク質を題材としてこれらの手法の計算精度を評価してきた。そこでR6年度課題においては、これらの計算技術を実践的な創薬テーマに適用し、推定した医薬品候補の結合ポーズ及び結合親和性を(1)共結晶構造における結合モード, (2)結合親和性の実測値(文献値), (3)本研究を通して実験的に測定する薬理活性,を用いて多角的に検証することで、計算コスト・計算精度の観点でこれらの手法を評価し、実用化に向けてシミュレーション・解析パラメータを精密化することを研究目的とする。

謝辞

- KBDDコンソーシアム参画企業
- バイオグリッドセンター関西 志水隆一、京都大学大学院 奥野恭史・荒木望嗣、横浜市立大学大学院 池口満徳・浴本亨、兵庫県立大学大学院 神谷成敏、大阪大学蛋白質研究所 Gert-Jan Bekker、筑波大学 広川貴次
- 本研究は、文部科学省ポスト「京」重点課題1「生体分子システムの機能制御による革新的創薬基盤の構築」および公益財団法人計算科学振興財団 研究教育拠点(COE)形成推進事業の協力を受けたものです。
- 本研究の結果は、理化学研究所のスーパーコンピュータ「京」「富岳」を利用するとともに、「京」「富岳」以外のHPCIシステム利用研究課題を遂行して得られたものです。(課題番号:hp120048、hp140042、hp150025、hp160010、hp170036、hp180011、hp190020、hp190021、hp200011、hp210048, hp220026, hp230025, hp240017)

2017~2020年:GUIの開発
(開発したアプリが簡単に使えるようにユーザインターフェイスを開発)



LINC-SHOWCASE 2024 Spring

WG04-PJ-MDAI：分子動力学AI

プロジェクト概要

分子動力学シミュレーションにより得られたトラジェクトリを、AIにより解析するツール（Giniツール）を利用し、その結果により得られた成果を参加メンバーで共有することで、利用方法の拡張やプログラムの向上を図る。

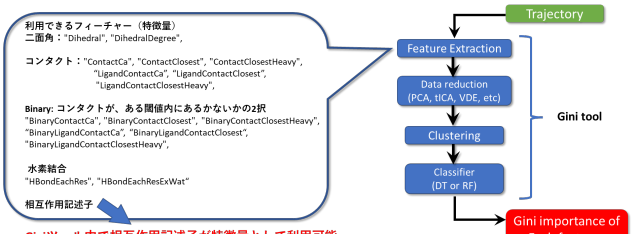
プロジェクト発足：2022年5月
期間：2年程度を予定
成果：
- LINC参加メンバーは、プロジェクト内で開発された機能を含むGiniツールを利用可能。
- 各社内での他の例へ適用可能なGiniツール利用のノウハウが取得できる
- 研究内容に関する新しい知識の共有
- Giniツールの利用方法の習得

Giniツールとは？

先のLINC PJ13で開発
分子動力学シミュレーション解析人工知能システム（GINIツール）
(J. Chem. Theory Comput. 2014, 10, 5217－5223)

- 各特徴量に対して分類の寄与度を数値化した、GINI importanceと呼ばれる値で、MDシミュレーションの構造をよりよく分類できる特徴量を数値化して検出。
- 人間では拾うことのできない特徴をうまく検出できる可能性もある。

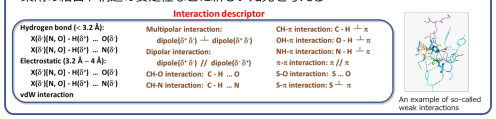
フローチャート (Gini ツール)



Giniツール内で相互作用記述子が特徴量として利用可能。

相互作用記述子（関連プロジェクト；WG04-PJ-IHBVS）

分子内・分子間で見られる以下に示すような弱い相互作用の数を検出することで、薬剤の結合や構造の安定性などに新しい知見を与える



活動内容

アカデミア課題（研究はアカデミアメンバーが行い、プロジェクト内でシェア）

1. 環状ペプチドの膜透過性へのginiツールの応用（山根（理研））
2. giniツールによるタンパク質-化合物複合体安定構造の自動抽出（荒木（京大））
3. giniツールによるキナーゼ活性化に重要な構造的特徴の自動抽出（荒木（京大））

LINCメンバー課題
個別研究課題（得られた成果は、プロジェクト参加者と共有）

1. PROTACのginiツールへの応用について（小野（田辺三菱））
2. そのほか希望する課題

共同作業課題（個別のテーマではなく、参加者で共同で行う）
1．新規記述子（相互作用記述子）の検証（Giniツールの公開に合わせて、開始予定）

共同作業提案（実施するかどうかは、プロジェクト化後検討）
1. プログラムの拡張
Cryptic Site検出ツールの作成や酵素デザインへの応用

Giniツールの勉強会・ユーザ会的な活動も行う
（参加メンバーがginiツールを使って解析を行えるようにする）
（giniツールの改良にかかわる意見をまとめる）

現在、Giniツールの利用方法のたたき台となる研究例を実施中

ALKとクリソチニブ複合体のGiniツールによる解析

Giniツールの公開

現在、プロジェクトに参加しているLINCメンバーにGiniツールを公開している

2022年11月5日に、FY2022バージョンを公開
2023年7月中に、FY2023バージョン公開予定

公開様式
Docker版
Tar-ball版
Singularity版（予定）

6月28日にハンズオンセミナーを実施（オンライン）

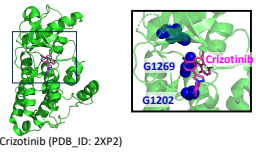
FY2023
トラジェクトリ版相互作用記述子の最新バージョンに対応

FY2022
新しい特徴量の追加
SASA：表面積
polsurf：極性表面積
LieEL：溶質溶媒間静電相互作用
LieVDW：溶質溶媒間VDW相互作用
MolVolume：溶質分子の体積

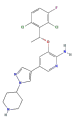
現在実施中の課題

ALKとクリソチニブ複合体のGiniツールによる解析

背景　ALKとクリソチニブ（Crizotinib）の複合体



ALK+Crizotinib (PDB_ID: 2XP2)



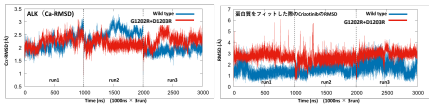
Crizotinib

- G1202R変異体はCrizotinibの阻害活性を減弱
- WTとG1202R+D1203RのMDシミュレーションの結果を解析
相互作用記述子とGINI解析を使って、MD trajectoryから自動的に新しい知識を引き出せるかを検討する。

方法
利用したMDデータ（京都大学　荒木先生提供）
ALK+クリソチニブ複合体（Wild type, G1202R+D1203R）
1μs × 3本（Total 300,000 frames)→30,000 frame

用いた特徴量：相互作用記述子（intDesc）
分類器：Random Forest
木の深さ：2、木の数：10

RMSDの比較（wild type vs G1202R+D1203R）



変異体のほうが、クリゾチニブの揺らぎが大きい

結果

特徴量順位の比較（5回の試行の結果、RF, depth=2）

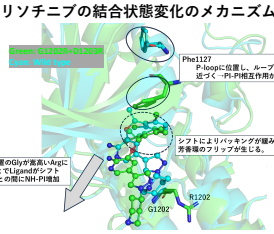
Rank	Run1	Run2	Run3	Run4	Run5
1	LIE9_OHPro	LIE9_OHPro	LIE9_OHPro	LIE9_OHPro	LIE9_OHPro
2	LIE5_OHPro	LIE5H_OHPro	LIE5H_OHPro	LIE5H_OHPro	LIE5H_OHPro
3	LIE5H_Hal_OHPro	LIE5H_OHPro	LIE9H_OHPro	LIE9H_OHPro	LIE9H_OHPro
4	LIE5H_OHPro	LIE5H_Hal_OHPro	LIE5H_OHPro	LIE5H_OHPro	LIE5H_OHPro
5	LIE9H_OHPro	LIE9H_OHPro	LIE5H_OHPro	LIE9H_OHPro	LIE9H_OHPro
6	LIE9H_Hal_OHPro	LIE9H_OHPro	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro	LIE5H_Hal_OHPro
7	LIE5H_OHPro	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro	LIE5H_Hal_OHPro	LIE9H_Hal_OHPro
8	LIEHal_Cl_OHPro	LIE9H_OHPro	LIE9H_Hal_OHPro	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro
9	LIE5H_Hal_OHPro	LIE5H_Hal_OHPro	LIE9H_OHPro	LIEHal_OHPro	LIE5H_OHPro
10	LIEvdwPro	LIE5H_Hal_OHPro	LIEHal_Cl_OHPro	LIE9H_Hal_OHPro	LIE9H_OHPro
11	LIE9H_Hal_OHPro	LIEvdwPro	LIE5H_OHPro	LIEvdwPro	LIEHal_Cl_OHPro
12	LIE5H_Hal_OHPro	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro	LIE5H_Hal_OHPro	LIE9H_Hal_OHPro
13	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro	LIE9H_Hal_OHPro	LIE9H_Hal_OHPro	LIE5H_Hal_OHPro
14	LIEHal_Cl_OHPro	LIEvdwPro	LIEvdwPro	LIEHal_Cl_OHPro	LIEEec_NH_OHPro
15	LIE5H_OHPro	LIE5H_OHPro	LIE5H_OHPro	LIE5H_OHPro	LIEHal_Cl_OHPro
16	LIEHal_Pl_OHPro	LIEHal_Cl_OHPro	LIEHal_Pl_OHPro	LIEHal_Pl_OHPro	LIEvdwPro
17	LIE9_OHPro	LIEHal_Pl_OHPro	LIE9_OHPro	LIE9_OHPro	LIE5H_OHPro
18	LIEHal_Br_OHPro	LIE9_OHPro	LIEHal_Br_OHPro	LIEHal_Br_OHPro	LIEHal_Pl_OHPro
19	LIEEec_NH_OHPro	LIEEec_NH_OHPro	LIEEec_NH_OHPro	LIEEec_NH_OHPro	LIEEec_NH_OHPro
20	LIE5H_OHPro	LIE5_OHPro	LIE5_OHPro	LIE5_OHPro	LIE9_OHPro
21	LIEQuoPro	LIEQuoPro	LIEQuoPro	LIEQuoPro	LIE5_OHPro

平均値

Feature ranking	Average ranking
1	1.0
2	2.0
3	3.0
4	4.0
5	5.0
6	6.0
7	7.0
8	8.0
9	9.0
10	10.0
11	11.0
12	12.0
13	13.0
14	14.0
15	15.0
16	16.0
17	17.0
18	18.0
19	19.0
20	20.0
21	21.0

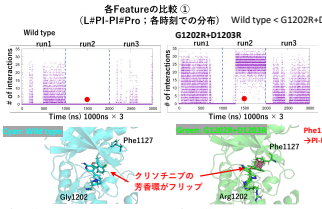
- 特徴量の順位が試行回数によらずほぼ一定であることを確認

クリソチニブの結合状態変化のメカニズム



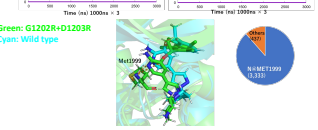
GINIツールにより自動的に注目すべき相互作用および残基が特定できた

各Featureの比較 ①
（L#PI-PI#Pro；各時刻での分布）　Wild type < G1202R+D1203R



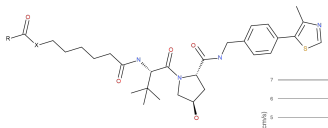
- 変異体のほうがPhe1127とのPI-PI相互作用を検出
- クリソチニブの芳香環がフリップすることが観測された

各Featureの比較 ⑥　Wild type << G1202R+D1203R
（L#NH-PI#Pro；各時刻での分布）



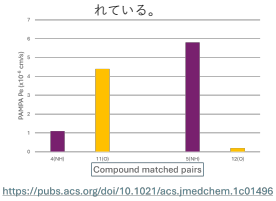
- 変異体はWTと比べて、NH-PIが多く観察された。
（その内訳の大半は、Crizotinibのピリジン環とMet1199のNHとの相互作用）
- クリソチニブがシフトすることが原因

PROTACモデル化合物の膜透過性 (クロロホルム溶媒中、水中のMcMD)



共通骨格を持ち、特定部分Rが共通で、Xの構造が異なる2組の分子構造（matched compound pair）において4種類のPROTACは、PAMPAアッセイより膜透過速度の違いが知られている。

R	X
	4(NH)
	11(O)
	5(NH)
	12(O)



https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c01496

上記4種類のPROTAC分子（4(NH), 11(O), 5(NH), 12(O)）について、水中およびクロロホルム溶媒中でマルチカノニカル分子動力学シミュレーション（McMD）を実施した。

本研究では得られた、トラジェクトリからリガンドー溶媒間についての相互作用記述子を特徴量としたGiniツールによりmatched compound pairを比較し、新しい知識が引き出せるかを検討する。

特徴量順位

4(NH) vs. 11(O)

CHCl3	round1	round2	round3	round4	round5
1	LIEOH_Hal_OHS	LIEOH_OHS	LIEOH_Hal_OHS	LIEOH_Hal_OHS	LIEOH_Hal_OHS
2	'LIEvdWHS'	LIEOH_Hal_OHS'	'LIEvdWHS'	LIHH_Hal_OHS	'LIEvdWHS'
3	LIEOH_OHS'	LIHH_Hal_OHS'	LIEOH_OHS'	'LIEvdWHS'	LIHH_Hal_OHS'
4	LIDpoSS	LIEOH_OHS	LIEOH_Hal_OHS'	LIEOH_OHS'	LIEOH_OHS
5	LIEHH_Hal_OHS'	LIEHal_Cl_OHS'	LIEHH_Hal_OHS'	LIDpoSS'	LIEOH_NHS
6	LIEHal_Cl_SHS	LIEOH_Hal_OHS'	LIEOH_NHS	LIEOH_SHS'	LIEOH_SHS
7	LIEHH_Hal_OHS'	LIEOH_NHS'	LIDpoSS'	LIEOH_Hal_OHS'	LIDpoSS'
8	LIHal_Cl_OHS'	LIHal_Cl_SHS'	LIHal_Cl_OHS'	LIHH_Hal_OHS'	LIEOH_Hal_OHS'
9	LIEOH_NHS'	LIDpoSS	LIEOH_SHS'	LIHal_Cl_OHS'	LIDpoSS'
10	LIEOH_NHS	LIEOH_SHS'	LIHal_Cl_SHS'	LIHal_Cl_SHS'	LIHal_Cl_SHS'

WAT	round1	round2	round3	round4	round5
1	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'
2	'LIEvdWHS'	'LIEOH_OHS'	'LIEvdWHS'	'LIEOH_OHS'	'LIHH_OH_OHS'
3	LIEOH_OHS'	LIHH_NH_OHS'	LIEEec_NH_OHS'	LIHH_OH_OHS'	LIEEec_OH_OHS'
4	LIEEec_NH_OHS'	LIHH_OH_OHS'	LIEEec_OH_OHS'	LIEEec_OH_OHS'	LIEOH_OHS'
5	LIEEec_OH_OHS'	LIEEec_NH_OHS'	LIHH_OH_OHS'	LIHB_NH_OHS'	LIEEec_NH_OHS'
6	LIHB_NH_OHS'	LIEEec_OH_OHS'	LIHB_NH_OHS'	LIEEec_NH_OHS'	LIHB_NH_OHS'
7	LIE_OHS'	LIE_OHS'	LIE_OHS'	LIE_OHS'	LIE_OHS'

CHCl3	round1	round2	round3	round4	round5
1	LIEOH_Hal_OHS'	LIEOH_Hal_OHS'	LIEOH_Hal_OHS'	LIEOH_Hal_OHS'	LIEOH_Hal_OHS'
2	'LIEOH_Hal_OHS'	'LIEvdWHS'	LIEOH_OHS	'LIEOH_Hal_OHS'	LIIDpoRS
3	LIEOH_OHS'	LIEOH_OHS'	LIEOH_OHS'	LIEOH_OHS'	LIEOH_SHS'
4	LIDpoRS'	LIEOH_NHS'	LIEOH_Hal_OHS'	LIHH_Hal_OHS'	'LIEvdWHS'
5	LIEOH_SHS'	LIEOH_Hal_OHS'	LIHH_Hal_OHS'	LIIDpoRS'	LIEOH_OHS'
6	LIHal_Cl_OHS'	LIHH_Hal_OHS'	LIEOH_NHS'	LIHal_Cl_OHS'	LIEOH_Cl_OHS'
7	LIHH_Hal_OHS'	LIHal_Cl_OHS'	LIDpoRS'	LIEOH_Hal_OHS'	LIEOH_NHS'
8	LIEOH_NHS'	LIDpoRS'	LIHal_Cl_OHS'	LIEOH_NHS'	LIEOH_Hal_OHS'
9	LIEOH_Hal_OHS'	LIEOH_SHS'	LIEOH_SHS'	LIEOH_SHS'	LIEOH_Cl_OHS'
10	LIHal_Cl_SHS'	LIHal_Cl_SHS'	LIHal_Cl_SHS'	LIHal_Cl_SHS'	LIHal_Cl_SHS'

WAT	round1	round2	round3	round4	round5
1	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'	'LIEvdWHS'
2	'LIHB_OH_OHS'	'LIEOH_OHS'	'LIHB_OH_OHS'	LIEEec_OH_OHS'	LIEOH_OHS'
3	LIEOH_OHS'	LIHB_OH_OHS'	LIEOH_OHS'	LIHB_OH_OHS'	LIHB_OH_OHS'
4	LIEEec_OH_OHS'	LIEEec_OH_OHS'	LIEEec_OH_OHS'	LIHB_NH_OHS'	LIEEec_NH_OHS'
5	LIHB_NH_OHS	LIEEec_NH_OHS'	LIHB_NH_OHS'	LIEEec_NH_OHS'	LIHB_NH_OHS'
6	LIEEec_NH_OHS'	LIHB_NH_OHS'	LIEEec_NH_OHS'	LIEOH_OHS'	LIE_OHS'
7	LIE_OHS'	LIE_OHS'	LIE_OHS'	LIE_OHS'	LIEEec_OH_OHS'

水中およびクロロホルム溶媒中でのGiniツールでの計算を各5回実施
- 特徴量の順位の収束性を確認

今後、PROTAC分子内部の相互作用を含む計算を検討

WG05

メディシナルケミストリー

WG05 では、これまで蓄積された膨大な創薬関連データに基づいて、オン・オフターゲット、ADMET 等のプロファイル予測 AI、合成難易度予測 AI、多目的同時最適化などの方法論を開発します。

【 WG リーダー 】
本間 光貴／理化学研究所

■ タスクフォースとプロジェクト

1 WG05-PJ-AgriAI　農薬　DB、予測モデルの構築など
2 WG05-tf-design　化合物デザイン　多項目同時最適化設計など
3 WG05-tf-model　モデル精度向上　追加学習、転移学習、能動学習、強化学習など
4 WG05-tf-synth　合成経路・容易性

G05 では、医薬品・農薬の設計に関する AIの情報収集、予備検討、手法開発、社会実装を進めている。設計を行うための AIとしては、大まかに予測 AIと生成 AIに分けることができる。

予測 AI には、タンパク質等のターゲットとの親和性予測、ADMET 予測、合成難易度予測、医薬品らしさ・特許性の予測等がある。一方、生成 AI としては、それらの予測 AI やドッキング等のシミュレーション結果を報酬スコアとして利用して、スコアが改善する構造を生成する AI が利用される。

LINC 第一期で GCN 記述子（kGCN, kMol）によるマルチタスク、マルチモーダル学習、MCTS と RNN による構成生成 AI である ChemTS の予備検討と手法開発を行った。WG05 によって生み出された技術は、AMED 次世代創薬 AI事業へと受け継がれ社会実装を進めている。

今年度の進捗として、製薬企業 17 社から 1500 万ポイント以上のデータの提供を受けてオンオフターゲット、ADMET プロファイルの予測精度が向上した。また、構造生成 AI については、AI 予測モデルの適用範囲を考慮した生成によって、医薬品候補の生成効率が向上することがわかった。

農薬の予測 AI については、WG05-PJ-AgriAI として活動を開始しており、農薬の環境影響を予測する AI システムの開発を目指している。

今年度は、農薬開発の早期段階での開発期間の短縮を実現するため、環境生物（ミジンコ、魚、ハチなど）の毒性予測 AI の開発と、環境中の分解性（特に土壌中の分解速度）のデータベース構築と予測 AI の開発を行った。

メンバーには、農薬・化学系 5 社、IT 系 1 社、アカデミア 3 機関が参加している。これまでに、水生生物ミジンコの急性毒性予測 AI のプロトタイプモデルを開発し、実用性の検証を行った。

また、欧州リスクレポート（RAR）から取得した 400 化合物の土壌中動態試験データを基に、153 化合物の分解指標とメタデータを抽出・整理したデータベースを構築した。今後は、これらのデータを用いた AI モデルの開発や、他の生物種の AI 開発を進める予定である。

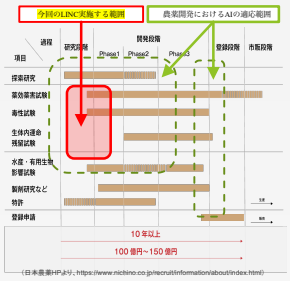
LINC-WG05-PJ-Agria 農薬の環境影響評価の予測

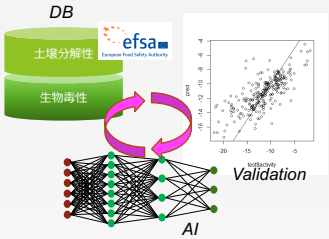
池田和由(理研・慶大)、米澤朋起(慶大)、小川真弘(クミアイ化学)、竹田至(クミアイ化学)、岩佐万実(日産化学)、星川優美子(日産化学)、飯野翔平(日本曹達)、松本健(日本曹達)、泰永涼子(日本農薬)、久間達彦(日本農薬)、大竹潤(日本農薬)、金岡怜志(三井化学CLS)、番場伸一(三井化学CLS)、新延弘章(三井化学CLS)、上島仁(ISP)、近藤圭(残農研)、福島勇二(残農研)、松本聡一郎(残農研) 、取出遥介(残農研)、大田雅照(理研)、本間光貴(理研)

目 的

農薬開発の早期（主に探索段階）において、期間短縮を実現するため、農薬の環境リスクを評価するAIシステムの開発を行う。

1. 環境生物（ミジンコ、魚、ハチ等）の毒性予測AIの開発を行う。
2. 環境中分解性（土壌中分解速度）のDB構築と予測AIの開発を行う。





ゴール

1. 環境生物毒性予測AIの開発
   - （ゴール）化学構造から各種環境生物への毒性を予測するAIモデルを開発する。
   - （使用用途）探索初期における化合物の安全性評価に活用する。
2. 環境中分解性のDB構築及び予測AIの開発
   - （ゴール）化学構造および土性情報から土壌中における化合物の分解速度を予測するAIを開発する。
   - （使用用途）探索初期における化合物の土壌残留性評価に活用する。最終的には、環境生物の毒性予測AIモデルと組み合わせることにより、環境リスクの評価ができるシステムを目指す。

結 果

I. 環境生物毒性：データ収集＆プロトタイプ検証

目的

- 環境生物の毒性予測AI (対象種：ミジンコ、魚類、藻類、ミツバチ)を構築するため、基盤となる各種毒性データを収集した。

進捗

- ECETOC AQUATIC TOXICITY(EAT) DATABASE、EnviroTox Database, EPA OPP Pesticide Ecotoxicity Database, ECOTOX Knowledgebaseより下記データを収集できた。

調査対象	収集総データ数	Endpoint	左記Endpointの総データ数	CASの重複なしデータ数
01 *Daphnia magna* acute toxicity	6,556	EC50	5,930	1,491
02 *Daphnia magna* chronic toxicity	4,543	NOEC	3,684	907
03 *Oncorhynchus mykiss* acute toxity	12,946	LC50	12,198	1,568
04 *Oncorhynchus mykiss* chronic toxity	4,367	NOEC	2,220	312
05 *Lepomis macrochirus* acute toxicity	9,833	LC50	9,573	1,335
06 *Lepomis macrochirus* chronic toxicity	237	NOEC	106	33
07 *Cyprinus carpio* acute toxicity	1,357	LC50	1,192	273
08 *Cyprinus carpio* chronic toxicity	749	NOEC	453	57
09 *Pimephales promelas* acute toxicity	9,601	LC50	8,565	1,415
10 *Pimephales promelas* chronic toxicity	5,521	NOEC	3,228	410
11 *Pseudokirchneriella subcapitata* toxicity	12,220	EC50	3,812	919
12 *Pseudokirchneriella subcapitata* toxicity	12,220	NOEC	2,170	855

II. ミジンコ急毒予測のプロトタイプAIの開発と検証①

目的

- 農薬・医薬品等の環境生物への影響を予測するため、ミジンコ急性毒性の予測AIプロトタイプを作成し、実用性を判断すべく、①外部データ、②各社データを用いて検証を実施した。

使用したデータ

- 半数遊泳阻害濃度EC50の単位をmol/Lに統一後、対数化したものを目的変数とした. 収集した6,556データから前処理（無機化合物・金属等の除去, 同一化合物データは外れ値除去後平均化）を実施し、得られた1,289データをプロトタイプ作成に用いた。

検証データ

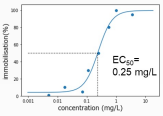
- 検証①：Furuhamaらのデータ（学習データと重複しない361データ)を用いた。 *SAR QSAR Environ Res. 2018 29, 725*
- 検証②：各農薬企業の内部データ（217データ）を用いた。

モデルの概要

- 学習アルゴリズム（RF, LightGBM, GBR)、使用記述子 (RDKit, Morgan Fingerprint), 10-fold cross validation

10-fold cross validation結果

学習アルゴリズム	使用記述子	MAE	MSE	RMSE	R2	RMSLE	MAPE
RF	RDKit + FP	0.796	1.179	1.082	0.583	0.197	0.204
RF	RDKit	0.801	1.182	1.084	0.582	0.198	0.206
LightGBM	RDKit + FP	0.778	1.139	1.064	0.595	0.193	0.196
LightGBM	RDKit	0.782	1.119	1.054	0.605	0.192	0.198
GBR	RDKit + FP	0.754	1.080	1.035	0.616	0.188	0.192
GBR	RDKit	0.755	1.103	1.046	0.609	0.189	0.191



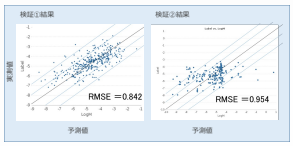
III. ミジンコ急毒予測のプロトタイプAIの開発と検証②

既存の予測モデルとの性能比較

モデル	r^2 of pesticides (56%)	r^2 of industrial chemicals (44%)	RMSE	r^2 for inside AD substances	r^2 for outside AD substances
ACD	0.43	0.39	**1.54**	0.55	0.24
ADMET Predctor™	0.51	0.56	**1.24**	0.60	0.44
DS TOPKAT	0.61	0.52	**1.20**	0.63	0.49
ECOSAR	0.31	0.46	**1.59**	0.37	0.16
Terra QSAR ™	0.51	0.52	**1.32**	0.57	n.a.
T.E.S.T.	0.4	0.47	**1.37**	0.44	n.a.
VEGA DEMETRA	0.73	0.37	**1.29**	0.62	0.62
VEAGA EPA	0.37	0.34	**1.54**	0.61	0.32
今回のプロトタイプ	-	-	**0.84**	-	-

検証①および②結果

予測誤差（絶対値）	検証結果	
	検証①	検証②
<= 1	277化合物 (77%)	129化合物 (59%)
1 <, <=2	68化合物 (19%)	55化合物 (26%)
2<	16化合物 (4%)	33化合物 (15%)



・既存モデルと比較して、同等以上の精度 (RMSE＝0.84)ができた。

・外部データに対しても、

59-77%の化合物のEC50 (mol/L)を実測値の0.1-10倍の範囲内で

85-96%の化合物のEC50 (mol/L)を実測値の0.01-100倍の範囲内で予測できた

今後の予定

- 各社データを用いた更なる検証、モデルの適用範囲を示せる指標の追加, モデルの改良等を目指す。
- 更に、魚類・藻類・ミツバチ急毒の予測モデルについても、順次、作成し、検証を進めていく。

IV. 環境中分解性：土壌中分解性データベースの構築

目的

- 化学構造および土壌特性から農薬の土壌中分解速度を予測し、地下水や表層水中における暴露濃度算定に応用することで簡易な生態リスクのスクリーニング評価を行う。

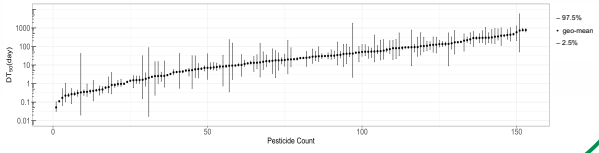
データの収集

- EUの農薬再評価レポート (RAR) から400化合物分の土壌中動態試験データを取得
- k-meansクラスタリング (類似度：Tanimoto係数, Dice係数) により，対象化合物を選抜
- 計153化合物の分解指標（DT_{50}, DT_{90}）およびメタデータ（土壌特性や試験設計)について約1080データポイントを収録したデータベースを構築

今後の予定

構築したデータベースをもとに，土壌中分解性を予測する機械学習モデルを構築する。

➡ 農薬毎に変動が大きいデータセットを上手くモデル化できるか？



まとめ

- ミジンコ急毒予測について、プロトタイプAIモデルを開発し、その精度検証を外部データ及び企業データを用いて行った。今後、各社データを用いた更なる検証、モデルの適用範囲を示せる指標の追加及びモデルの改良等を行う。
- 他の生物種（魚類・藻類・ミツバチ）のデータ収集を行った。また魚類急毒プロトタイプAI開発に着手し、今後そのモデル検証を行う予定である。
- 環境中分解性予測について、土壌中分解性データベースの構築を行った。構築したデータベースをもとに、土壌中分解性を予測する機械学習モデルを開発する予定である。

WG06

製剤・マテリアルインフォマティクス

WG06 では、研究者の経験や勘に大きく依存してきた医薬品の処方設計・材料開発や素材探索を多目的最適化・解釈可能 AI などの IT 技術を活用することで、設計プロセスを形式知化することを目標としています。

【 WG リーダー 】
岩田 浩明／鳥取大学 医学部

■ タスクフォースとプロジェクト

1 WG06-PJ-molxtl　　結晶予測 AI
2 WG06-PJ-FormDevAI　　製剤設計 AI（休止中）

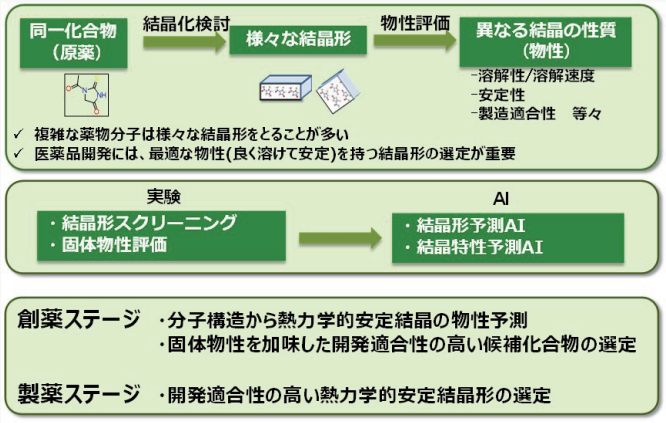
■ WG06-PJ-molxtl

高い付加価値を持つ結晶性物質の獲得を目指し、結晶構造を予測する機械学習モデルの構築に向けて、以下の取り組みを進めています。

1. 分子構造から適切な結晶構造を予測するシミュレーション方法と機械学習、AI の導入

2. AED を使用した結晶エネルギー算出の機械学習法の改良
（以前の LINC22a の成果を活用して複雑な分子にも適用）

3. 自由エネルギー差計算を用いた高精度な結晶算出のための機械学習モデルの構築
結晶形 AIから得られる結晶構造情報を活用し、結晶特性の予測に役立つシミュレーション法と機械学習モデルの開発を進めています。

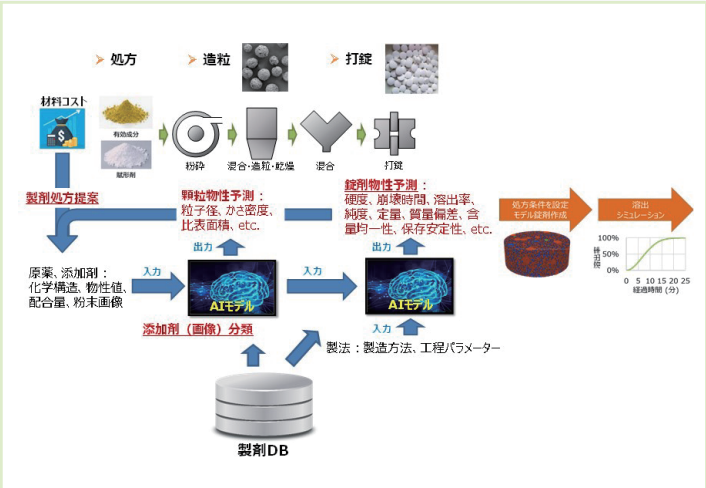


■ WG06-PJ-FormDevAI

国立研究開発法人 新エネルギー・産業技術総合開発機構（NEDO）からの委託を受け、「新薬開発を効率化・加速する製剤処方設計 AI の開発」プロジェクトを推進しました。

具体的には、効率的な学習データの生成・構築技術、マルチモーダル・マルチタスク型の機械学習アルゴリズムと能動学習フレームの開発、解釈可能な AI モデルによる暗黙知の可視化技術を開発し、医薬品開発の加速に貢献する製剤処方設計 AI を構築しました。

これらの要素技術の応用先となる研究開発パートナーを募集しています。一緒に研究開発を検討していただける方は、ぜひご連絡ください。



WG07

ADMET・トランスレーショナルリサーチ

WG07 では、薬物のヒト in vivo における体内動態や薬効 / 毒性発現の時空間的かつ定量的な予測を実現化するために、創薬過程で評価される in vitro データや文献情報の有効活用を支援するシステム開発に取り組むことを目標とします。

【 WG リーダー 】
前田 和哉／北里大学

■ タスクフォースとプロジェクト

1	WG07-PJ-PathoDiag	AI 駆動型画像解析 / 病理診断の自動化プラットフォームの開発
2	WG07-PJ-QSPAI	Quantitative Systems Pharmacology と人工知能を用いた臨床薬効予測基盤の構築
3	WG07-tf04	薬の構造からの主クリアランス経路予測の高度化、 薬物動態パラメータの予測に対する深層学習導入効果の評価（リーダー独自提案）

たとえ薬効が前臨床研究において強い医薬品候補化合物が探索されたとしても、ヒトに適用できる医薬品になるには、大きなギャップがあります。

その主な要因の一つとして、その化合物のヒトにおける薬物の体内動態特性 (ADME properties) の悪さや、予期せぬ毒性発現 (Toxicological effect) が挙げられます。これら ADMET 特性の予測が薬効の予測と並走しない限りはヒトに適用可能な医薬品の創製はできません。

さらにヒト臨床においては、薬効や ADMET 特性一つとっても個体間変動が生じることから、それらの指標の平均値の予測のみならず、ばらつきの定量的予測もまた求められます。その際に、分子メカニズムを忠実に再現するような数理モデルを介した機構論的な定量的解析が必要となります。

本 WG では、様々な前臨床データを活用して、ヒト臨床での薬効・副作用の定量的予測を行うことでトランスレーショナルリサーチを推進する技術開発を志向しています。現行では、次の2つのプロジェクトが進行中です。

それ以外でも、WG リーダーは、特に in vitro 実験の結果に基づく定量的モデル解析を介したヒト臨床予測には豊富な経験を有しておりますので、新規テーマをご提案いただければ、プロジェクト立ち上げのお手伝いをさせていただきます。

■ AI 駆動型の自動化病理診断

動物の毒性試験等に由来する病理組織画像は、これまでは病理医がマニュアルで1枚1枚判読を実施しており、多大な労力がかかっていました。また、画像取得法も会社により細かいところが異なることから、異なる会社由来のデータをまとめて解析することは困難でありました。
本プロジェクトでは、複数の由来の異なる動物の毒性試験由来病理画像の判読の自動化を目指した新規 AI 技術を開発しています。

AI + QSP 解析による
■ 臨床での薬効の定量的予測

薬効 / 副作用に関わる分子の挙動と医薬品の体内動態を支配する異物解毒系分子の挙動の両方を考慮した上で、ヒトにおける医薬品の薬効の経時変化を定量的に予測する方法論として、定量的システムズ薬理学 (QSP) の考え方に基づく数理モデル解析が脚光を浴びていますが、その方略には未だ課題も山積しています。
本プロジェクトでは、複雑な QSP モデルにおいて、臨床での薬効 / 副作用のばらつき予測のために必須な、現実的なパラメータ範囲を有する" virtual human" を効率よく自動生成可能な AI の開発をしています。

WG08

治験・市販後・メディカルアフェアーズ

WG08 では、医薬品等の治験・臨床研究に関する情報、承認審査及び製造販売後に得られる情報を活用し、戦略的、かつ効率的な治験実施を可能とし、また製造販売後の医療安全対策に寄与する AI 術開発を目指しています。

【WG リーダー】
岡本 里香／神戸医療産業都市推進機構 医療イノベーション推進センター (TRI)
中津井 雅彦／山口大学
安倍 理加／理化学研究所

■ タスクフォースとプロジェクト

1	WG08-PJ-HERO	HEOR：費用対効果分析	
2	WG08-PJ-CONNECT	臨床試験患者リクルーティング効率化	
3	WG08-TF03 （PV/PMS:Pharmacovisilance）準備中	PVにおける医療情報を用いた AI 活用のモデル検討	
4	WG08-TF05 (MI:Medical Information）準備中	医薬品情報活用のための TF	①医薬品情報の構造化・標準化（WG00 との連携） ②構造化・標準化データ活用検討

■WG08-PJ-HERO

電子カルテ情報によるリアルワールドデータ（RWD）を用いた病態進行を予測する AI の開発

東 美恵 1、西田陽介 2、野末 卓 2、黒須真介 3 、大場史記 3、阪田幸則 1、安倍理加 4、長谷川亜樹 5、岡本里香 6、中上裕有樹 7、中津井雅彦 8
1. エーザイ、2. NTT データ、3. EA ファーマ、4. 理研、5. 京大、6. 神戸医療産業都市推進機構 TRI、7.下関市大、8. 山口大

レセプトデータとカルテ（診療録）データを統合した「千年カルテ」の2次利用が、平成30年5月1日施行の次世代医療基盤法により可能となった。

臨床試験という特殊環境下ではなく、「千年カルテ」を用いた実臨床データ（RWD）を活用し、HTAに必要な生存関数の算出とその検証に挑戦する

がん領域であれば、無増悪から増悪（再発）に進行する時点、更に増悪から死亡に至る時点を抽出することにより、全生存期間（OS）無増悪生存期間（PFS）の算出、生存曲線を作成し、先行研究と比較し外挿可能性を図る

■WG08-PJ-CONNECT

臨床試験患者リクルーティング効率化

岡本里香 *1, 中津井雅彦 *2, 安倍理加 *3, 佐々木栄二 *4, 星野美由紀 *4 , 越田愛深 *4, 朝生祐介 *5, 大場史記 *6, 角野忠宏 *7, 中上裕有樹 *8, 伊藤眞里 *9, 深川明子 *9, 藤田靖之 *1, 鍵村達夫 *1, 東出智嗣 *1, 中川聡史 *1, 岡本匡史 *1
*1 神戸医療産業都市推進機構 TRI,*2 山口大学 ,*3 理化学研究所 ,*4 日本アイ・ビー・エム株式会社 ,*5 第一三共株式会社 ,*6EA ファーマ株式会社 ,*7 協和キリン株式会社 ,*8 下関市立大学 ,*9 医薬基盤・健康・栄養研究所

【背景】臨床試験の約 80% が症例登録目標数の組み入れを設定したスケジュール内に達成できておらず、その後のプロセスの遅れに繋がっている *1 。

*1 出典：薬事日報 (2017) 被験者リクルートメントの現状と課題

【ゴール】臨床試験遅延の原因である「患者リクルーティング」において、AI 技術を用いることで、患者・医療機関・製薬企業にとって使いやすく、効率的なプラットフォームの構築を目指す。

■WG08-TF03 (PV/PMS:Pharmacovisilance)

準備中

PV における医療情報を用いた AI 活用のモデル検討

岡本里香 神戸医療産業都市推進機構 TRI

【目的】
❶ PV分野における AI活用のためのガイダンス作成の基盤となる知見の提供、❷ 副作用症例の評価、情報整理、規制当局における迅速な施策立案の推進

■WG08-TF05 (MI:Medical Information)

準備中

医薬品情報整備とその活用検討

安倍理加　理化学研究所

【目的】
医薬品情報整備（WG00/08 共同）と標準化データ活用の社会実装をイメージ検討
（CMC エクスメディカ、ロゼッタ）

WG08-pj-CONNECT
臨床試験患者リクルーティング効率化

先崎心智[*1],佐々木栄二[*1],星野美由紀[*1],越田愛深[*1],朝生祐介[*2],大場史記[*3],角野忠宏[*4],
岡本里香[*5,6],中津井雅彦[*7],安倍理加[*8],中上裕有樹[*9],伊藤眞里[*10],深川明子[*11],藤田靖之[*6],鍵村達夫[*6],東出智嗣[*6],中川聡史[*6],岡本匡史[*6]

[*1]日本アイ・ビー・エム株式会社,[*2]第一三共株式会社,[*3]EAファーマ株式会社,[*4]協和キリン株式会社,
[*5]京都大学,[*6]神戸医療産業都市推進機構医療イノベーション推進センター,[*7]山口大学,[*8]理化学研究所,[*9]下関市立大学,[*10]大阪大学薬学研究科,[*11]医薬基盤・健康・栄養研究所

WG08-tf09-pj-CONNECT　プロジェクトゴール

臨床試験遅延の原因である「患者リクルーティング」において、先端技術を用いることで、患者・医療機関・製薬企業にとって使いやすく効率的なプラットフォームの構築を目指す

「患者リクルーティング」に注力するのは、既存のデジタルプラットフォームが十分に有効活用されておらず、未だアナログ依存のプロセスが多いためにデジタル活用によって改善が見込めることが理由である

「患者リクルーティング」のプラットフォーム構築を通じ，これまで分断されていた「患者の治験情報へのアクセス」と「治験情報」を連携させ，全国の治験と患者から被験者リクルーティングを実施できる「3方良し」の環境を実現する

患者

患者が治験情報を適切に入手し、能動的に治験参加を選べる環境を作る

患者は治験参加の機会を増やせる

製薬企業

治験期間を短縮させる。
また、治験期間延長・治験中止・終了のケースを減少させる

「3方良し」の環境を実現

医療機関

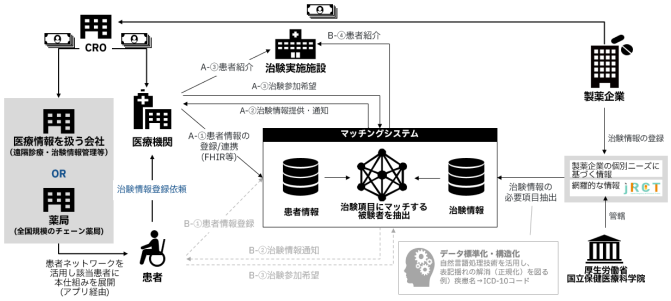
クリニック、かかりつけ医：
患者に対して、良質な医療を提供し、中長期で患者と信頼関係を構築する

治験実施病院：事業機会を拡大する

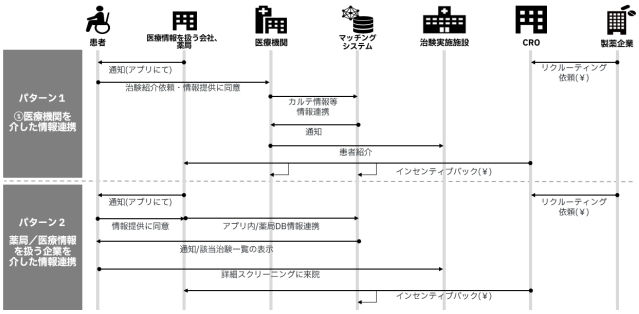
目指す姿

患者データ×jRCT情報　マッチングシステム　相関図

jRCT上に公開されている情報と 患者データとのマッチングを行うことで、網羅的な治験情報から潜在的な被験者を効率的に抽出し、リクルーティングの効率化を実現する。



より多くの患者にリーチするためのアプローチ：プロセス詳細



目指す姿に関するヒアリング結果

治験データ（jRCT活用）

- 厚生労働省として、**jRCTの改善や外部との連携を促進する動きはないように見える。**患者リクルーティングプラットフォームとして目指す姿は**jRCTへの集約ではなく、民間企業による情報をかみ砕いた形での情報提供**である。
- **製薬企業が出せる情報は既にjRCTに公開されている為、jRCT情報を活用するのが良い。**治験届など追加詳細情報を収集することは、製薬企業の知財戦略面から難しい。**製薬企業はメリット、または法的義務が無い限り、能動的にはjRCT記載情報以上の公開は難しい。**
- jRCT記載の以下の項目を活用し、登録基準に合う患者を抽出できるのではないか。
  - 第一優先：対象疾患名と薬剤使用歴、年齢、性別
  - 第二優先：過去の罹患期間、有効性評価を含む検査値、既往歴、合併症

  （※「主たる選択基準・主たる除外基準」はjRCT登録必須項目であるが記載方式や粒度感が治験毎に異なる。）

患者データ

病院からの情報提供：

- PHRを基軸とした場合と比較し、データ信頼性が高い。
- 一方、治験リクルーティングにおいて、リアルタイム患者情報が重要であるが、医療機関が保有している情報(レセプト情報や電子カルテ情報) は後ろ向きが多い。
- **データ連携に伴う倫理審査プロセスに関して懸念**あり。以下を整理する必要がある。
  - 匿名化した状態でデータ収集するか
  - 治験案件単位で実施する仕組みか
  - 公共性が高い運用とするか、営利的な運用とするか
  - 治験倫理審査委員/通常の倫理審査委員どちらにどのタイミングで申請するか

患者からの情報提供：

- **治験に少しでも興味のある患者から真正性の高いデータを集める**ことを目指すと良い。（例：パネル検査のゲノム診断結果など）
  - 参加希望がない患者のデータを大量に持つよりも、治験に少しでも興味のある患者のデータをどのように集めるかにフォーカスしたほうが良いのではないか。
- **治験へ関心がある新たな患者が、自ら治験情報へアクセス・応募できることは価値がある。**
  - 患者が情報提供する場合情報粒度は荒くなるが、n数が多い治験においては、ニーズがあるのではないか。
- マイナンバーの保険証利用が進み、患者の権限で詳細の医療情報共有が出来る形が良い。**将来マイナンバーとの連携を想定してシステム設計すると良い。**

薬局活用

- 薬局は処方情報のみ保有しており、疾患名等は登録されていない。**不足情報をどのように補うかが課題**となる。
- リクルーティングに薬局を活用する場合、**以下の様な薬局と相性が良い**のではないか。
  - 人的リソースに余裕がある薬局
  - 患者とリレーションが築けている薬局 （例：薬剤師が定着し移動が少ない、地域の方が継続して来られている、訪問服薬指導を実施している）
- 以下3点が満たされる場合、**薬局・薬剤師の協力が得られることは十分考えられる。**
  - ①金銭的インセンティブ　②既存の薬剤師の業務に取り込めるようなスキーム　③薬局による治験リクルーティング促進の意義
- **医師会や患者からの反発は特に心配ない。**医師会は、金銭的インセンティブが減額とならない場合、特に反発はされない。また、患者は、薬剤師側が適している患者をある程度判断できる可能性が高い。

リクルーティング市場調査

国内リクルーティングプラットフォーム調査

- 業界として、**治験情報の網羅的な検索ができ、応募まで完了するプラットフォームはない。**
  - これまでの患者リクルートは、Web上の治験プラットフォームを効率的に活用できていないのではないか。
  - 既存の医療機関・患者ネットワークの活用（医療機関内の広告、紹介、PROのデータベースなど）を通じて、独自の患者リクルーティングが主流となっているのではないか。

	民間プラットフォーム	公共プラットフォーム
掲載情報	● 広告掲載が主流であるため、**掲載数は限定的** ● 患者への情報発信に焦点を当て、患者が理解しやすい表現で、掲載情報は最低限に留めている	● **網羅的**な情報がそろっている ● 記載内容はユーザーにとって理解しづらい
検索条件	● フリーワード検索や、年齢等必要最低限の項目	● 民間と同様の項目に加え、研究や試験の内容に関する詳細情報を含む （例）試験名や実施組織、計画の説明、アウトカム評価項目等（民間では項目なし、または情報掲載のみ）
機能	● Web上で簡易的に参加可否の判断がつく**スクリーニング機能を持つサービスが多い。**ただし、**jRCT情報から転載された治験情報はスクリーニング出来ないため、**スクリーニング可能な治験数は限定的	● 保存や比較機能あり ● **応募やスクリーニング機能は不足**

海外動向調査

日米共通

- 米国でも、**治験リクルーティングの一元化はされておらず、患者リクルーティングは課題**とされている。
  - 治験情報は、国立公衆衛生研究所(NIH) と米国医薬食品局(FDA) が米国国立医学図書館(NLM) を通じて共同で提供するClinical.govに集約されているが、これは患者リクルーティングを目的としておらず、ポータルなども存在しない。
- **Web利用でのリクルーティングの主要なプラットフォームは存在しておらず、日米で大きな差異は見受けられない。**

日米差異（米国特徴）

- **米国における医療機関と患者にとって、治験に参加する際の動機付けは日本より高い**と考えられる。
  - 大学や病院だけでなくクリニックも積極的に治験に参加しており、被験者獲得と臨床試験の実施が収入獲得手段の一つとなっている。
  - 多数の無保険者や民間保険による受診可能な医療機関の限定を受ける市民が存在するため、米国では医療費の優遇が治験参加者にとって大きなインセンティブになっていると考えられている。（厚労省資料より）
- 米国対面でのリクルーティングは、2020年以降、**大規模ドラッグチェーンが参入**し、薬局ネットワークやヘルスケアサービス会員にリクルーティングを行っている。
  - コロナ治験で緊急大量の治験参加者リクルーティングが課題となったことから全米1位のドラッグストアチェーンであるCVS Healthに次いで2位のWalgreenも参入したものの、CVS Healthは2024年での撤退を表明した。
  - 米国では、クリニック併設のかかりつけ薬局が広く浸透していることが影響していると考えられる。薬局・クリニックを有するため患者リクルーティングから臨床試験の実施まで一貫して行うことが可能。

WG08 PJ-HERO

電子カルテ情報によるリアルワールドデータ（RWD）を用いた病態進行を予測するAIの開発

東 美恵[1]、西田陽介[2]、野末 卓[2]、井上裕文[2]、小林慎[3]、大場史記[4]、阪田幸則[1]、安倍理加[5]、長谷川亜樹[6]、岡本里香[6]、中上裕有樹[7]、中津井雅彦[8]
1.エーザイ、2. NTTデータ、3.クレコンメディカルアセスメント、4. EAファーマ、5.理研、6.京大、7.下関市大、8.山口大

- **2023年度の活動実績**
  4-7月：千年カルテデータから算出する生存曲線と比較するための文献候補を収集し整理し、最終的に最も適した1文献*を決定した。
  8月-11月：千年カルテ利用目的等委員会へ申請し、データ利用の承認を取得した。
  12月-2月：千年カルテからデータ抽出のための患者条件を検討し、解析データセットを抽出した。
  3月：解析データをもとに生存曲線作成を開始した。
- **見出された課題**
  アウトカムの一つである無増悪生存期間（PFS）を算出するための「増悪」を、千年カルテデータにおいてどのように定義するか？

【背景】

- 日本では、2019年4月より医薬品・医療機器に関し費用対効果の制度導入が開始された。
- 費用対効果の推計のために分析対象技術と比較対照技術を用いた場合の病態進展モデルが構築されるが、特に抗がん剤の評価では生存関数の推計が重要な作業の一つとなる。
- 分析対象技術の生存関数の推計は臨床試験結果を用いて推計することが多いが、比較対照技術は臨床試験に含まれていないこともあり、別途推計の必要がある。
- 市販のレセプトデータは医療費の推計に用いられることもあるが、生存関数の推計のためには必要な情報が十分でないケースがある。
- 生存関数の推計方法には一長一短があるため、複数の選択肢が存在することでお互いの短所を補いより適切な生存関数の推計が可能となることが期待される。
- 千年カルテは複数の医療機関のカルテデータを蓄積しており、生存関数の推計に必要な情報も含まれていることから、本研究ではこれを用いた生存関数の推計を試みる
- 最終的にはAIを用いて、主要な背景データをインプットすることで生存関数が自動的に推計されることを目標とするが、千年カルテからの生存曲線の描画可能性の確認を本件の第一段階の目的とする

【研究の全体像】

レセプトデータとカルテ（診療録）データを統合した「千年カルテ」の2次利用が、平成30年5月1日施行の次世代医療基盤法により可能となった。

↓

「千年カルテ」を用いた実臨床データ（RWD）を活用し、費用効果分析に必要な生存関数の算出とその検証に挑戦する

↓

がん領域を対象として、無増悪から増悪（再発）に進行する時点、更に増悪から死亡に至る時点を抽出することにより、全生存期間（OS）、　無増悪生存期間（PFS）の算出、生存曲線を作成する。

【方法】

1. 千年カルテからの生存曲線の作成に必要なデータの抽出：

奏効、進行などの転帰情報を正確に抽出するため、テキスト記述に対しては自然言語処理手法を活用し、電子カルテのテキスト領域の臨床サマリー、経過記録を基にルールベースでの表現辞書の作成、またはAIを活用した機械学習による表現辞書を作成する。表現辞書の作成に必要な自然言語処理手法及びAI技術の同定も行う。次に表現辞書に基づき、千年カルテから転帰情報を効率的に抽出する仕組みを構築する*2。

2. 解析対象患者の設定と生存曲線の作成：

抽出した転帰情報を特定の患者属性ごとに推計するために必要となる学習用サンプルと検証用サンプルを設定する。次に、これらのサンプルを用いて患者属性ごとの各種転帰を推計し、生存曲線を描画する方法および描画された生存曲線に対して理論分布を適合する方法を構築する。

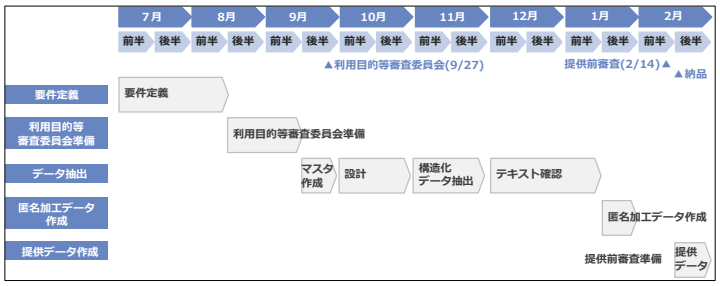
3. 生存曲線作成の精度向上：

実施した生存曲線の描画や理論分布の適合について、学習用・検証用サンプル間において、まずは統計理論的に同等の生存曲線かどうか比較検討を行う。次に、項目ごとのデータの特性や分布を反映する補間手法やBootstrapなどのサンプリング手法を適用し、注目する時間軸での生存率や生存分布を推定することで、モデルの精度向上およびその限界を検討する。

【研究スケジュール】

科研費（基盤C）	2023年度		2024年度		2025年度	
	前半	後半	前半	後半	前半	後半
EMRからの生存曲線の作成に必要なデータ抽出	→					
解析対象患者の設定と生存曲線の作成		→	→			
生存曲線の精度向上			→	→	→	
過去手法との比較検証					→	→

【2023年度の作業工程】



【増悪の定義】

PFS算出における増悪イベント発生の定義は、パターン①だけでよいか？

パターン	Nライン	N+1ライン
パターン① アテゾリズマブ単剤治療中に増悪	アテゾリズマブ単剤治療 ★増悪	次のレジメン
パターン② アテゾリズマブ単剤治療の切り替え後に増悪	アテゾリズマブ単剤治療	次のレジメン ★増悪
パターン③ 増悪の記載は見られないがレジメンの切り替えがある	アテゾリズマブ単剤治療	次のレジメン

【千年カルテからの抽出項目一覧】

no	カテゴリ	項目	データの内容	データ型	データ構造
1	患者情報	患者ID	仮ID	文字列	構造化
2		年齢	アテゾリズマブ初回投与時の年齢（10歳刻み）	文字列	構造化
3		性別	1:男性,2女性	文字列	構造化
4		PS	パフォーマンスステータス0~4	文字列	非構造化
5		PS相対日	PSの記載日 – アテゾリズマブの初回投与日	数値	非構造化
6		喫煙歴	1:喫煙歴あり,0:喫煙歴なし	文字列	構造化
7		BMI	BMI：数値	数値	構造化
8	アウトカム	PFS	アテゾリズマブ投与～増悪または打ち切りまでの期間	整数	非構造化
		増悪パターン	1:アテゾリズマブ単剤治療期間中,2:アテゾリズマブ単剤治療終了後	文字列	構造化
		アテゾリズマブ投与期間	アテゾリズマブ単剤治療の初回終了日－アテゾリズマブ単剤治療の初回投与日	整数	構造化
		後続レジメンまでの期間	アテゾリズマブ単剤治療の次のレジメンの初回投与日－アテゾリズマブ単剤治療の初回投与日	整数	構造化
9		PFS打ち切りフラグ	1:打ち切りあり,0:打ち切りなし	文字列	非構造化
10		OS	アテゾリズマブ投与～死亡または打ち切りまでの期間	整数	構造化
11		OS打ち切りフラグ	1:打ち切りあり,0:打ち切りなし	文字列	構造化
12	疾患情報	ステージ	ステージ:1~4	文字列	構造化/非構造化
13		ステージ相対日	ステージの記載日 – アテゾリズマブの初回投与日	数値	非構造化
14		組織学	1:扁平上皮癌,0:その他（腺がんまたは大細胞がんまたは不明）	文字列	構造化
15		EGFR	1:変異あり,0:変異なし	文字列	非構造化
16		PDL1	PD-L1 TPSの数値（%）	数値	非構造化
17		PDL1相対日	PDL1の記載日 – アテゾリズマブの初回投与日	数値	非構造化
18		肝臓転移	1:転移あり,0:変異なし	文字列	構造化
19		脳転移	1:転移あり,0:変異なし	文字列	構造化
20		手術歴	1:手術歴あり,0:手術歴なし	文字列	構造化
21		放射線治療	1:放射線治療あり,0:放射線治療なし	文字列	構造化
22	薬剤	免疫関連副作用	1:免疫関連副作用あり,0:免疫関連副作用なし	文字列	構造化
23		治療ライン	治療ライン数(2nd,3rd,≧4th)	文字列	構造化
		後続治療ライン	1:後続の治療ラインあり、0:後続の治療ラインなし		
24	検査	Alb	Alb(g/dL)	数値	構造化
25		Alb_フラグ	1:Alb>3.8g/dL 0:Alb≦3.8g/dL	文字列	構造化
26		CRP	CRP(mg/dL)	数値	構造化
27		CRP_フラグ	1:CRP>0.89mg/dL 0:CRP≦0.89mg/dL	文字列	構造化
28		LDH	LDH(U/L)	数値	構造化
29		LDH_フラグ	1:LDH>227U/L 0:LDH≦227U/L	文字列	構造化
30		好中球数	好中球数(/mm3)	数値	構造化
31		好中球数_フラグ	1:好中球数>4500/mm3 0:好中球数≦4500/mm3	文字列	構造化
32		リンパ球数	リンパ球数(/mm3)	数値	構造化
33		リンパ球数_フラグ	1:リンパ球数>1000/mm3 0:リンパ球数≦1000/mm3	文字列	構造化
34		単球数	単球数(/mm3)	数値	構造化
35		単球数_フラグ	1:単球数>500/mm3 0:単球数≦500/mm3	文字列	構造化
36		血小板数	血小板数(/mm3)	数値	構造化
37		血小板数_フラグ	1:血小板数>250,000/mm3 0:血小板数≦250,000/mm3	文字列	構造化
38		NLR	好中球/リンパ球	数値	構造化
39		NLR_フラグ	1:NLR>5.0 0:NLR≦5.0	文字列	構造化
40		LMR	リンパ球/単球	数値	構造化
41		LMR_フラグ	1:LMR>1.5 0:LMR≦1.5	文字列	構造化
42		PLR	血小板/リンパ球	数値	構造化
43		PLR_フラグ	1:PLR>262 0:PLR≦262	文字列	構造化
44		WBC	WBC(mm)	数値	構造化
45		変換フラグ	1:変換あり　0:変換なし（変換対象：好中球数,リンパ球数,採求数）	文字列	構造化

References;
1. Katayama Y, et al. Significance of inflammatory indexes in atezolizumab monotherapy outcomes in previously treated non-small-cell lung cancer patients. Sci Rep. 2020;10(1):17495.
2. Araki K, et al. Developing Artificial Intelligence Models for Extracting Oncologic Outcomes from Japanese Electronic Health Records. Adv Ther. 2023;40(3):934-950.

WG09

人材育成

WG09 では、LINC での人材育成を実施します。あるべき人材のペルソナイメージを抽出し、イメージを実現するために必要な講座、OJT や意見交換会などを実施いたします。特にマネジメント層、経営層の DX リテラシー取得に注力いたします。

【 主なメンバー 】（順不同）

中田一人（日本電気）
上島　仁（システム計画研究）
中嶋久士（エクサウイザーズ）
谷村直樹（みずほリサーチ & テクノロジーズ）
渡辺悠太（スキルアップ NeXt）
兵庫淳志（第一三共）
今泉　明（味の素）
大里直哉（日本農薬）
小島健介（協和キリン）
＜事務局＞
小柳智義（京都大学 / 筑波大学）
志水隆一（都市活力研究所）

当初、人材育成ワーキング（WG09）では、IT スキルの習得を中心に知識獲得型の人材育成事業を検討していましたが、知識獲得型の人材育成は LINC の会員ですでに事業として提供しており、ここは既存の企業に任せるべきと WG09 は判断しました。

そこで、LINC では業界の DX 化に対応できる人材像をイメージし、そのイメージに合致した人材育成のメニューを検討してきました。現在は、人材像としての「ペルソナイメージ」を作成し終えたところで、今後これらの人材育成を実現するために具体的になにをするかを検討するの段階に来ております。

LINC ならではの人材育成として、DX リテラシーの向上やマインドセットの醸成などを主目的に、経営層、マネジメント層のワークショップの開催、サロン的な意見交換会の開催、他社事例の紹介などのアイデアが上がっており検討段階です。DX が一般的になってきた中で、WG09 では今後の DX 推進を担う人材育成に協力いただける方を広く募集しております。

【 ペルソナイメージ 】

	事業責任者／リーダー	事業遂行者／メンバー
先駆者／能動的	• 社内データを有効活用したいが、関係組織に対して、どのようにその有効性の説明したら納得してもらえるかを悩んでいる. • デザイン思考などの知識とスキルを有しているが、最新技術の本質を捕らえきれていないため、組織のビジネス変革のデザインまでに至っていない. ビジネス変革と同じ手法が研究開発の変革に、そのまま適用できるのかを知りたい. • メンバは高度な専門知識とスキルを持っており、IT系の研修も計画的に推進している. 体系的な知識の習得が有効と考えており、現業務に負担がかからないよう、効率のよい受講プロセスを知りたいと思っている. • 技術全体の俯瞰をもとに課題抽出と解決提案ができる現場メンバを増やしたいと考えているが、それに適した研修がなく、技術単位でのスキル習得を目的とした研修ばかりであった.	• 研修を計画的に受講しており、担当業務への具体的な適用をしたいと考えているが、よいアイデアが浮かばない. • データ活用のアイデアを持っているが、リーダーを説得できるロジックを組めない. • プログラミング知識とスキルを持ち、最新技術の情報収集に努めているが、担当業務への最適な手法の選択や条件の検討を試みたところ、知識不足を感じた. • ITスキルをもっているため、リーダーからDX推進を命じられたが、必要となるIT技術の範囲が広いため、どこから手を付ければよいか迷っている. • 現在の業務プロセスを変えたいと考え、研修を受講したいが、この組織の風土では言いにくい状況である. • 現場に課題があり、IT技術を活用して解決したいと考えているが、どこから手を付ければ良いのか・どんな技術が適しているのか・何を勉強したら良いのか・誰に聞けば良いのか分からない. 気軽に相談できる人が、敷居の高い別部署ではなく、身近にいてほしい. • 上位職から部署のDX推進の指示があり、 不足しているITスキルを習得しようとしているが、学ぶべき分野が多岐に渡り、どこから手を付ければよいか迷っている. また、各メンバのスキル（初級、中級、上級を目指す人）レベルに合わせた育成計画を立てるため、適切な教材を探している.
追随者／受動的	• 失敗したときのインパクトを最小限に抑えるために、体系的な知識＆スキルの習得と、社内外の先行事例の収集＆分析から始めようとしている. • 経営層からの推進指示に対して、自部門でメンバを含めて検討したが、アイデアが出なかった. • 自部門メンバの専門性は高く、先端的であり、他部門連携も問題がないため、DXのメリットを感じていない. • 全社的なWGが優秀な人材を集めて立ち上がっているので、自部門で検討し提案することは不要と判断し、そこから落ちてくるのを待つことにした. • 社内事例には成功例だけでなく失敗例や重複する例もあり、責任を取れる旗振り役が必要と考えるが、行動に至らない. • 上位層から現場のデジタル化推進の指示があり、メンバへニーズを聞いたが、なかなか提案が上がらなかった. その原因が、現場研究者のITリテラシーの低さ、あるいはWet実験のみへの興味にあると考えている.	• 業務に追われ、研修受講などの時間を取れない. • 自主的な研修受講では評価されないため、上司の指示に従わざるをえない. • 自ら調査する時間がないので、成功事例の中から担当業務へそのまま適用可能なものだけを知りたい. • いろいろな情報を目にするが、担当業務への適用を想像できないので、その先へ進んでいない. • 他部門所有のデータと連携する場合、多くの関係者との調整が必要なため躊躇しており、研修受講のモチベーションが低下している. • 上位職からデジタル化や効率化の推進指示があった. しかし、生物試料を扱う実験では、毎回の手技手法が異なり、調整の必要もあり、サンプル間誤差も考慮しているため、デジタル化への移行が難しく提案に至っていない.