

日本の「AI・データ駆動型創薬研究」を俯瞰する

柚木克之，山西芳裕

はじめに

機械学習の技術革新に端を発した人工知能（AI）の進歩は科学史に残る革命となりそうである。例えば、深層学習は高次元かつ大規模なデータから潜在的構造を抽出することを可能とし、現実世界の複雑性を対象とする科学研究の方法論に変革をもたらしている。この変革の流れは分子生命医科学や創薬科学にも、AI駆動型研究として波及しつつある。異なる実験条件や測定系から得られる高次元データを一括して学習させる解析の枠組みが多数提案され、生体システムの多面的理解を志向する研究が進展している。一方、AI駆動型研究は大量のデータを必要とする。この1点において、AI駆動型研究はもう1つの潮流として存在していたオミクスなど大規模データ駆動型研究と必然的に合流する。

本書の編者2人は、どちらかというとながら木がマルチオミクス・トランスオミクスを中心とするデータ駆動型、山西がバイオインフォマティクス・ケモインフォマティクスを中心とするAI駆動型の研究を展開してきた。その過程で創薬や薬理作用を研究対象とするに至ったのだが、2人とも薬学部や薬学系の大学院の卒業生ではない。そのようなわれわれ2人が、2025年3月18日に「AI駆動型・データ駆動型の創薬と薬理研究」と題したシンポジウムを日本薬理学会年會にて開催した。このシンポジウムには芦田広樹（Eight Roads Ventures Japan, 当時）、Frank Pun（Insilico Medicine Hong Kong Limited）の両氏を講演者として招き、学術機関・企業を問わず多くの方々にご来場いただいた。講演後の議論も盛り上がり、このテーマに対する創薬関係者の関心の高さを垣間見ることができた。

さらに、このシンポジウムが本増刊号の企画につながった。当日会場にいた「実験医学」編集部の山口恭平氏より、1カ月後の4月18日には企画の打診を受けることとなった。そこで編者となったわれわれ2人は、わが国の「AI・データ駆動型創薬研究」を俯瞰できるような1冊をめざして、関連分野の研究者に可能な限り声をかけて執筆をお願いした。残念ながら今回はご都合が合わずに寄稿がかなわなかった方もおられるし、筆者陣が固まってから「しまった、この方を忘れていた！」とわれわれの不注意でお声掛けできなかった研究者の方もいる。したがって、当然ながら本邦の「AI・データ駆動型創薬研究」の総力は本書がカバーするよりもずっと大きいのだが、少なくともそれを知るための足がかりにはなっているであろう。

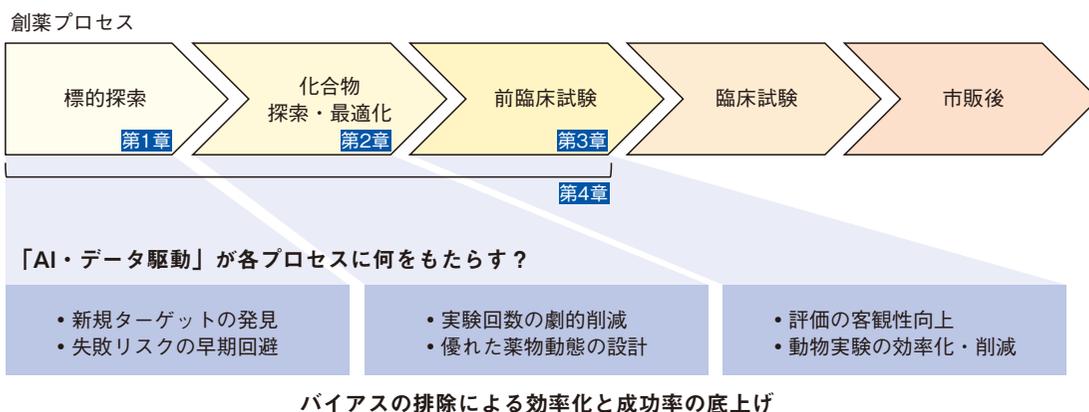


図1 「AI・データ駆動」が創薬プロセスに何をもたらすか

1. 本書のニーズ駆動的な読み方

本書は冒頭から順に読む必要はない。本書を手にとってくださった読者諸氏にはそれぞれ創薬ステージ（図1）に応じたニーズがあることと思うので、各自のニーズに直結する記事を最初に読み、次に同じ章の他の記事へと広げていくのがいいだろう。

全4章のうち、第1章は創薬標的分子に関する次のようなニーズに応えることをめざす。

- ・創薬標的タンパク質を網羅的に探索・同定する手法を知りたい
- ・標的分子周辺で、薬剤に応答するパスウェイ・ネットワークの時空間的挙動を追跡したい

第2章では、医薬品候補分子に関する次のようなニーズに応える。

- ・創薬標的タンパク質に作用する化合物を探索したい
- ・医薬品候補化合物の構造を生成したい
- ・ヒット化合物からリード化合物へ構造を最適化したい

第3章も引き続き医薬品や候補分子についてのセクションだが、主に作用機序の解明に重点を置き、以下のようなニーズに応える。

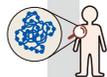
- ・いわゆる「昔の薬」の作用機序を解明して新規標的の探索に応用したい
- ・表現型スクリーニングで浮上した候補化合物の作用機序を解明したい
- ・単独の分子を狙うだけではなく、分子ネットワーク全体を医薬品で制御したい

第4章はAI創薬に関する産学の取り組みに関して、以下のようなニーズに応える。

- ・広く公開されているAI創薬の情報リソースについて知り、研究に役立てたい
- ・製薬企業の研究職を志望しており、企業におけるAI創薬の取り組みについて知りたい

第1章

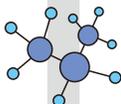
創薬標的の探索



- プロテオームデータを活用 [1章-1](#) [1章-2](#) [1章-3](#)
- メタボロームデータを活用 [1章-4](#)
- イメージングデータを活用 [1章-5](#)
- オミクスなどの高次元データ × 機械学習 [1章-6](#) [1章-7](#) [1章-8](#)

第2章

医薬品候補分子の探索と最適化



I. 探索研究

- 化合物・タンパク質間の相互作用や結合部位の予測 [2章-1](#)
- 化合物ライブラリから化合物探索・知識統合・推論 [2章-2](#)
- 既存薬からの探索 [2章-3](#)

II. 生成研究



【低分子化合物】

- 天然物も活用した潜在空間の構築と探索 [2章-4](#)
- 多目的最適化分子の設計AI [2章-5](#)
- 化学反応データを活用した構造生成AI [2章-6](#)
- Hit-to-Leadの統合支援AI [2章-7](#)
- 化合物の画像キャプション技術による生成AI [2章-8](#)
- トランスクリプトームデータを活用した構造生成AI [2章-9](#)

【中分子・高分子】

- mRNAワクチン設計への応用 [2章-10](#)
- がんワクチン開発への応用 [2章-11](#)
- タンパク質言語モデルの応用 [2章-12](#)

第3章

医薬品候補分子の作用機序の解明と評価



I. 作用機序，メカニズム解明

- 空間トランスクリプトームによる測定系 [3章-1](#)
- 作用機序解明事例（メトホルミン） [3章-2](#)
- マルチオミクスデータによる健康評価 [3章-3](#)

II. 候補分子の評価

- 作用する生物学的パスウェイの同定 [3章-4](#)
- 効能と副作用の予測 [3章-5](#)
- 薬物動態予測 [3章-6](#)
- 毒性評価 [3章-7](#)

第4章

国内外の動向



I. 国内のAI創薬プロジェクトの概要と展望

- AGIS（2024年度～，理研） [4章-1](#)
- DAIIA（2020年度～2024年度，AMED） [4章-2](#)
- PRISM（2018年度～2022年度，内閣府） [4章-3](#)
- AI-SHIPS（2017年度～2021年度，経済産業省） [4章-4](#)

II. 製薬会社の創薬におけるAI活用の取り組み

- 第一三共 [4章-5](#)
- 小野薬品工業 [4章-6](#)
- アステラス製薬 [4章-7](#)
- 中外製薬 [4章-8](#)
- グローバル投資家の視点から [4章-9](#)

図2 本書の内容

2. 各章のねらいと内容

それでは各章の内容を概観しよう（[図2](#)）。タイトルと要旨の中間のような内容なので、読む記事が決まっている方は読み飛ばしていただいて差し支えない。

1) 創薬標的の探索（第1章）

第1章は生体組織中の疾患治療の標的となる生体分子を探索するための戦略を紹介する。プロテオーム計測を応用した標的探索技術について3篇，メタボローム解析による探索について1篇，イメージングデータを用いた技術について1篇，オミクスも含めた高次元データに機械学習ベースの手法を適用して創薬標的分子やリポジショニング候補薬を探索する志向の記事を3

篇ご寄稿いただいている。

プロテオーム計測による標的探索技術のなかでは、まず熱を利用した標的同定手法として細胞サーマルシフトアッセイと二次元電気泳動を組み合わせた2DE-CETSA（第1章-1：室井・川谷・真田・堂前・長田）、質量分析を用いる熱プロテオームプロファイリング（TPP）（第1章-2：幡野・松本）を紹介する。いずれも、タンパク質の熱変性温度がリガンドとの結合によって変化する現象を利用した手法である。2DE-CETSAとTPPはプロテオームワイドな標的探索を可能にするが、これに加えて薬剤の結合部位に関する情報も併せて得られるのがLimited Proteolysis-Mass Spectrometry（LiP-MS）（第1章-3：小形・石濱）である。LiP-MSでは、プロテアーゼで限定消化した試料を質量分析に供する。結果得られたペプチド切断パターンから標的タンパク質の構造変化を検出し、これを立体構造にマップすることで化合物の結合部位を推定できるという発想に基づいている。

さらに、表現型に近いオミクス階層であるメタボロームから疾患マーカーを探索する方法についても紹介する（第1章-4：三枝）。

イメージングデータを用いる手法としてはProtein Localization and Modification-based Covariation Network（PLOM-CON）法（第1章-5：村田・加納）を紹介する。PLOM-CON法は、蛍光抗体染色を用いてタンパク質の量や局在の変動を時系列データとして取得し、変動が同調する分子群の「共変動ネットワーク」を同定する画像解析技術である。薬剤投与による共変動ネットワークの再配線など、動的な時空間的応答を精密に追跡できる技術である。

高次元データに機械学習ベースの手法を適用する探索手法については、まず都築らの手法を紹介する（第1章-6：都築・岡田・丸山・金・小澤・柚木）。「過去に蓄積された全文献、オミクスデータ、転写制御ネットワークの構造などを入力として、次に実験すべき分子群を出力する」という「逐次的分子選択問題」を定式化し、これを複数の探索エージェントによる群探索を用いて解決するフレームワークを提案する。

創薬標的分子の探索手法として、ゲノム情報・オミクス情報・臨床情報を駆使した手法を紹介する。既存の創薬標的分子を別の疾患に再利用するためのターゲットリポジショニングや、ゲノムワイド関連解析GWASとトランスクリプトームワイド関連解析TWASの融合により創薬標的分子を予測する機械学習手法が提案されている（第1章-7：難波・山西）。また、医薬品有害事象報告情報や化合物-タンパク質間相互作用情報を用いて治療標的分子を探索する方法が提案されている（第1章-8：酒井）。実際に乾癬に応用し、同定した治療標的分子候補の有用性を疾患モデルマウスで実証している。

2) 医薬品候補分子の探索と最適化（第2章）

第1章が標的サイドのニーズに応える事例紹介だったのに対し、第2章では医薬品候補分子サイドのニーズに焦点を当てた記事を集めた。

I. 探索研究

疾患の創薬標的タンパク質が決まっている場合は、その標的タンパク質を制御（阻害または活性化）し、治療効果につながる化合物の探索が実施される。

まず、ヒット化合物探索やリード化合物開発の工程を効率化するため、標的タンパク質と候補化合物の相互作用や結合部位を予測する深層学習やシミュレーションと組み合わせる方法論の近年の動向について解説する（第2章-1：富井）。化合物の探索では、化合物ライブラリが大きくなると計算コストが大きくなるという問題に直面する。効率的かつ高精度に化合物探索を行うために有用な、巨大データベースで高速検索を可能にする技術jXBW、知識統合の精度向上技術RAS、推論精度向上技術StructRAGを紹介する（第2章-2：田部井）。化合物探索から知識統合、推論支援まで一貫した創薬支援システム構築の展望を解説する。

いち早く必要な治療薬を臨床現場に届けるためには、安全性が確認されている既存薬のなかから対象疾患の治療薬候補を探索するドラッグリポジショニングやドラッグリパーバシングが有用である。ここでは、疾患・薬剤・遺伝子からなる「知識グラフ」に基づき、任意の疾患に対する治療薬候補を予測する多層型創薬プラットフォーム RePhaIND[®]を紹介する（第2章-3：中山）。ケーススタディとしてCOVID-19へのリパーバシング候補薬の同定の例を紹介する。

II. 生成研究

近年の生成AIの発展により、医薬品候補を既存の分子から探索するだけでなく、医薬品候補となる分子を新規に生成・設計する研究がさかんになってきている。対象の化合物も、低分子から中分子、高分子まで広がっている。

所望の物性や活性をもつ化合物を生成するさまざまな深層生成モデルを解説する。天然物と合成物の橋渡しを行うため、条件付き生成や大規模学習により物性分布を再現し、未踏領域から合理的な新規骨格の提案などの応用を実現する研究を紹介する（第2章-4：榊原）。また、分子構造と性質の関係を探索しながら、複数の所望の性質を満たす新たな分子を創出する生成AIの技術について解説し、分子生成フレームワーク「ChemT5」による分子設計の原理と応用事例を紹介する（第2章-5：藤井・寺山）。ヒット化合物やリード化合物の化学構造を生成する深層生成モデルは、創薬研究の現場に広く浸透しつつあるが、合成可能性への考慮が十分とは言えない。ここではリアクトームデータを活用し、合成可能性を考慮して医薬品候補化合物を生成する深層生成モデルに関する近年の研究動向やその応用例について概説する（第2章-6：森本・高田・山西・津田）。ヒット化合物をリード化合物に磨いていくHit-to-Lead工程では、活性・物性・合成容易性など相反する要件の同時最適化が求められる。従来の「実験して確認する」手法から「予測して設計する」アプローチの技術を解説し、構造的妥当性や相互作用の制御、合成経路の自動提案、多目的最適化を実現する最新のAI群を紹介する（第2章-7：関嶋）。Molecular Topographic Mapによる化学構造の画像化と画像キャプション技術をもとにした化学構造の生成AIのしくみについて解説し、特定の創薬標的タンパク質の阻害剤の最適化戦

略について紹介する（第2章-8：吉森）。

分子生物学における測定技術の発展と生命科学ビッグデータの整備に伴い、ケミカル情報だけでなくオミクスデータなど生命情報の創薬応用も試みられている。新たなオミクス創薬アプローチとして近年注目されているトランスクリプトームに基づく構造生成AIについて解説し、遺伝子摂動応答トランスクリプトーム情報を利用して任意の標的タンパク質に対する阻害剤・活性化剤を生成する構造生成AIについて紹介する（第2章-9：松清・山中・山西）。

深層生成モデルは、低分子だけでなく、中分子や高分子の生成にも応用できる。ベイズ最適化や能動学習などの数理的探索法、AIが仮説を提案し実験も主導する反復最適化ループ、タンパク質言語モデルによる配列空間探索の技術を解説し、mRNAワクチンの設計への応用を紹介する（第2章-10：伊莉・川上）。がんワクチン開発のボトルネックとなっていた、がん特異的腫瘍抗原ペプチド（ネオアンチゲン）と主要組織適合遺伝子複合体（MHCクラスII）との結合親和性予測を高精度に実現する深層学習モデル「MTL4MHC2」と、その開発に用いた「マルチタスク学習」について概説する（第2章-11：一久・二階堂）。医薬品の機能・構造予測、配列生成など特定のタスクに効率的に特化させるタンパク質言語モデルを解説し、抗菌ペプチド探索や抗体配列設計などのさまざまな応用を紹介する（第2章-12：大谷・藤原・清水）。

3) 医薬品候補分子の作用機序の解明と評価（第3章）

第3章は表現型スクリーニングなどでヒットした候補化合物や、いわゆる「昔の薬」の作用機序をAI・データ駆動的に解明することをめざした事例の紹介記事を集めた。

I. 作用機序，メカニズム解明

トランスクリプトーム計測技術，特に1細胞レベルのトランスクリプトームや空間トランスクリプトームの最新動向を紹介し，がんにおける薬物応答をプロファイリングする研究への応用にも触れる（第3章-1：芳賀・鈴木絢子・鈴木 穰）。また、「昔の薬」のなかには，作用機序についての論争がいまだに決着していないものがある。2型糖尿病薬メトホルミンの作用機序解明にトランスオミクス解析を応用した事例を紹介する（第3章-2：幡野・柚木）。オミクスデータは疾患をシステムの状態として定量評価する目的でも活用できる。渡邊らは肥満のマルチオミクスデータ（ヒト血液，腸内細菌叢由来）を機械学習・システム生物学を用いて解析し，従来の指標では検出できなかった健康状態変数を定量評価した。これにより，分子標的からネットワーク全体へと視野を広げた「システム標的」とその制御を提唱する（第3章-3：渡邊）。

II. 候補分子の評価

薬物応答の遺伝子発現データは，薬物の潜在的な標的タンパク質，作用パスウェイ，新規効能の予測に活用されてきた。この際，データ解析上の妨げとなってきたのが欠損値や未観測値である。ここではテンソル分解に基づいて欠損値を補完する手法TIGERSについて解説し，細胞タイプを区別して薬物の作用機序同定に応用した事例を紹介する（第3章-4：岩田）。

2024年のノーベル化学賞の対象にもなったAlphaFoldにより、ゲノムワイドに明らかにされたタンパク質の立体構造情報は、創薬研究にとっても非常に有用なリソースである。ここでは、ヒト全タンパク質の立体構造情報とすべての既存薬物の全ペアについてドッキングシミュレーションを行い、各薬物のポリファーマコロジーを考慮しながら、潜在的な効能と副作用を網羅的に予測する情報技術を紹介する（第3章-5：澤田・坂尻）。

薬物動態（吸収・分布・代謝・排泄）や毒性の評価は、金銭・時間の両面で高コストな工程である。さまざまな機械学習手法を用いて薬物動態を予測することで、特性評価のコスト圧縮をめざした事例を紹介する（第3章-6：江崎）。また、毒性評価における動物実験を減らしていこうという世界的な潮流を踏まえて、インシリコの代替手法や毒性試験データベースを紹介する（第3章-7：竹下）。

4) 国内外の動向（第4章）

いわゆるAI創薬と関連する国内の大型プロジェクトの紹介を4篇、企業の動向について5篇ご寄稿いただいた。「Ⅰ. 国内のAI創薬プロジェクトの概要と展望」では理化学研究所の科学研究基盤モデル開発プログラムであるAGIS（第4章-1：泰地）、産学連携により構築された創薬AIプラットフォームDAIIA（第4章-2：本間）、官民研究開発投資拡大プログラムPRISMの支援により構築された疾患統合データベースと関連AI開発（第4章-3：夏目）、AIを活用した毒性予測システムAI-SHIPS（第4章-4：船津）を紹介する。

「Ⅱ. 製薬会社の創薬におけるAI活用の取り組み」では、企業でのAI活用について第一三共（第4章-5：芹沢）、小野薬品工業（第4章-6：江頭）、アステラス製薬（第4章-7：森）、中外製薬（第4章-8：寺本）の実例をそれぞれ紹介する。また、投資家の視点から概観したAI創薬のマクロな動向についても寄稿していただいた（第4章-9：芦田・鈴木）。各社、それぞれ出せる情報には限りがあったことは想像に難くないが、可能な範囲内で事例をご紹介いただいた。製薬企業の研究職を志望する方には特に一読を推奨したい章である。

本書は薬をめぐる異分野融合の触媒である

冒頭でも述べたように本書は「AI駆動型・データ駆動型」というキーワードとかわりのある創薬研究を幅広く紹介することをめざして企画された。結果として出来上がった目次の執筆陣を見ると、生物学系（薬学・医学など）とそれ以外（主として情報・数物系）の出身者が半々くらいになった。これはとてもよいことだと考えている。科学史家のトーマス・クーンは、ある科学分野にイノベーションを起こすのは若い人が分野外からやってきた人である、という意味のことを言っている。生物学系と異分野の研究者が、どちらが上とか、偉いとかそういった関係にはならず、対等の協力関係で議論することで、どちらか一方のみでは考えもしなかったアイデアが生まれ続けるのだ。そしてその後にはきっと薬学・医学と情報・数物系の両分野の

素養をもつ融合型人材が頭角を現してくるのだろう。将来、そんな優れた人材のなかに、本書がその進路選択のきっかけであったという方が1人でもいれば、本書の目的は果たされたと言っている。

〈筆者プロフィール〉

柚木克之：理化学研究所生命医科学研究センター チームディレクター。2004年慶應義塾大学大学院政策・メディア研究科博士課程 単位取得満期退学。'05年博士（学術）取得（指導教員：富田 勝教授）。'06～'10年慶應義塾大学理工学部 助教。'10～'13年東京大学大学院理学系研究科 特任助教。'13～'16年同助教。'15～'19年科学技術振興機構さきがけ「疾患代謝」領域 研究者（兼任）。'17～'20年理化学研究所統合生命医科学研究センター Young Chief Investigator。'20年より現所属。海王星は数理モデルによって存在が予言され、ほぼ予言通りの位置に発見された惑星です。私も統合オミクス解析を起点に予測生命科学の方法論を構築し、分子生命科学にとっての「海王星の発見」に至りたいと考えています。

山西芳裕：名古屋大学大学院情報学研究科複雑系科学専攻生命情報論講座 教授。2005年京都大学大学院理学研究科博士課程修了，博士（理学）。'05～'06年フランス École Nationale Supérieure des Mines ポスドク。'06～'07年京都大学化学研究所 特任助手・特任助教。'08～'12年フランス Curie Institute & Mines ParisTech 常勤研究員。'12～'18年九州大学高等研究院・生体防御医学研究所 准教授。'15～'19年科学技術振興機構さきがけ「社会情報基盤」領域 研究者（兼任）。'18～'23年九州工業大学大学院情報工学研究院生命化学情報工学研究系 教授。'23年より現所属。'25年より愛知県がんセンター研究所異分野融合研究開発分野 分野長（兼任）。バイオインフォマティクスやケモインフォマティクスの研究に取り組んでいます。どんな疾患に対しても治療標的や医薬品分子を提案できる AI技術の開発をめざしています。