

2.1.4 AI創薬・インシリコ創薬

(1) 研究開発領域の定義

創薬において、コンピュータを用いて医薬品開発を支援する領域である。シミュレーションと AI（ビッグデータ解析・機械学習）を活用し、プロセスの自動・高速化、コストダウン、および大量情報処理による知識の抽出を経て創薬に貢献する。理論ベースの量子シミュレーションと、データから学習させる機械学習の二つの方法が存在する。量子シミュレーションでは、正確性は担保されるが膨大な計算コストが必要である。機械学習では、比較的計算コストが低い精度に難がある。両者は目的に応じて使い分けられている。

(2) キーワード

ヴァーチャルスクリーニング、リード最適化、ADMET、ドラッグリポジショニング、人工知能、機械学習、Deep Learning、ビッグデータ解析、バイオインフォマティクス、ケモインフォマティクス、シミュレーション、分子動力学計算（MD）

(3) 研究開発領域の概要

[本領域の意義]

創薬のプロセスは、薬剤標的の探索、リード化合物の探索・最適化を経て、臨床段階へと至る多岐に渡る専門領域が連結した工程であり、創薬の現場ではこの工程に沿って試行錯誤やフィードバックを繰り返している。医薬品の候補となる化合物を探索、絞り込む際の要素技術には、化合物の収集・管理、アッセイ系の構築・実施、評価データの利用などがある。近年の創薬においては、例えば「がん患者の幹細胞の mRNA 変異を分子生物学的に解析して見つかった分子標的に対して、調製した標的タンパクを用いて生化学者がアッセイ系を構築し、化学合成した化合物の活性を測定する。活性の最適化の為には標的タンパクと化合物の共結晶の X 線構造解析を用いて結合様式を見出し、計算科学によって更なる活性及び選択性の向上を図る。」といったことが行われている。さらには試験管内や培養細胞において生物活性や選択性がいかに優れているとも生体内での薬効や安全性が確認されなければ医薬品に不適であるので、薬物代謝等の考察から生体内での安定性を最適化し、薬理的に薬効を確認する必要がある。

しかし、十分な分子プロファイリング（薬理薬効機序を理解するための化合物標的とその表現型パスウェイ情報）が得られていない医薬品が7～8割に及ぶと言われる。患者ヘルスレコードの電子化や、ゲノムをはじめとするオミクス情報の拡充などにより、健康医療ビッグデータは急速に拡大している。その結果、医薬品の研究開発は標的分子主導型からデータ主導型に大きく転換しつつある。さらに、Precision Medicine への流れにより、予防や患者への適切な医療の提供が要求されている。

近年、製薬産業を取り巻く状況が大きく変化している。低分子を中心とした創薬ターゲットが枯渇傾向にあるいわれであり、既存の研究手法の延長線上で新薬創出の成功確率が低下していることは大きな課題である。開発コストも年々上昇し、製薬産業のコスト体質を悪化させている。さらに先進国では、高騰する薬剤費を含む社会保障費を背景に薬価抑制圧力が強く、日本では、イノベーティブな新薬として登場した抗がん剤 Nivolumab (Opdivo®) の薬価が短期間に約 75 % 引き下げられる等、製薬会社は従来のビジネスモデルを再考せざるをえない状況である。

深層学習によって飛躍的な進歩を遂げた AI は近年あらゆる分野で目覚ましい成果を上げている。2018年4月にFDAは、医師の手を借りず眼底画像から糖尿病性網膜症の発症を診断する IDx Technologies 社の

IDx-DRに自律型AI診断システムとして世界初の承認を与えた。AI診断システムは、2027年までに650億ドルを超える規模の市場に成長するとの見込みもある。その他にも、強化学習・自然言語解析等のAIにおけるあらゆる研究分野が製薬産業の研究開発プロセスに適用可能であることから、AI基盤技術である機械学習の手法を創薬に応用しようという試みがある。これまでも、サポートベクターマシンに代表されるカーネル法、lassoなどのスパースモデリング、ランダムフォレスト、ディープラーニングなどの機械学習の手法は既に創薬工程の様々な問題（化合物の活性予測、化合物の標的予測、化合物の薬理作用・副作用予測など）で提案されてきている。また、この機を受けてIT企業がAI技術を用いて創薬事業へ参入する動きもあり、創薬産業全体に大きな波が押し寄せている。

[研究開発の動向]

2020年、Exscientia社のAI創薬プラットフォームCentaur Chemist™を使って創出したセロトニン5-HT1A受容体アゴニストDSP-1181について、共同研究を進めていた大日本住友製薬社が強迫性障害を適応症として日本で臨床試験を開始した。また、NEC社のAI技術を使って設計されたがんワクチンTG4050が、卵巣がんおよび頭頸部がんを対象に臨床試験を開始した。Deep Genomics社は、プラットフォームとするAI workbenchを使って同定したウィルソン病の治療標的となる遺伝子変異に対する核酸医薬DG12P1を、2021年中に臨床試験入りさせる。DSP-1181の例では、通常4.5年程度掛かる創薬プロセスを1年未満で完了、DG12P1は、標的探索から18ヶ月で開発候補化合物の同定を終えたとされており、AIの活用による創薬プロセスの効率化はいよいよ実用段階に近づいている。創薬ベンチャーにおいては、その40%が化合物ライブラリからのスクリーニングに、28%が創薬標的の探索に、17%がde novoドラッグデザインに、それぞれAIを活用していると報告されており、様々なデータの取得・収集の深化、計算機性能の向上、計算手法の進展とともに、計算科学的手法を創薬に用いることが定着しつつある。

1980年代には定量的構造活性相関（QSAR：Quantitative Structure-Activity Relationship）という、化合物の物理化学的性状や結合エネルギーを基にして計算科学で化学構造と生物活性の関係を予測する手法が主としてアカデミアで盛んに行われたが、製薬業界では定着しなかった。また1990年代にはコンビナトリアルケミストリーという、固相法を用いて非常に多くの化合物を半ば無差別的に合成し、活性の高い化合物をその中から探し出すという手法が注目を浴びた。世界中の大手製薬会社がこぞってこの手法を取り入れ、また多くの製薬ベンチャー企業がコンビナトリアルケミストリーの技術を売り物にして設立された。しかしながらこの手法も注目に値するような結果をもたらすことなく、2000年代初頭までには製薬企業で創薬の為に使われることはなくなった。それに代わるものとしてFBDD（Fragment-based drug discovery）という、活性はそれ程高くない分子量200程度の化合物を複数個つなぎ合わせる事によって活性の高い分子を作り上げるという手法が提唱されてきた。

現在、in silico創薬研究で開発されている技術は以下の5つが主なものである。

- ・ 医薬品の標的となる疾患原因分子（主にタンパク質）の探索技術
- ・ 薬理活性化合物を探索する技術（ヴァーチャルスクリーニング技術）
- ・ 活性化合物を最適化する技術（リード最適化技術）
- ・ 吸収、分布、代謝、排泄、毒性（ADMET）や薬物動態を予測する技術
- ・ 既存薬の新しい適応可能疾患を予測するドラッグリポジショニング技術

ここで、これらをカバーする学術分野はバイオインフォマティクス、ケモインフォマティクス、計算化学などであり、これらの分野は創薬応用を主目的として研究開発されてきたわけではなく、各分野で研究開発される

計算技術の応用例として創薬に適用されることがほとんどである。

上記5つの技術に分けて詳細を説明する。

・医薬品の標的となる疾患原因分子（主にタンパク質）の探索技術

疾患サンプルと正常サンプルの分子レベルや分子ネットワークレベルの違いを計算によって解析し、疾患発症・進行の分子メカニズムの解明と原因分子を同定することを目的とするものである。これまで実施されてきた研究は、疾患の分子メカニズムの解明などの医学研究に付随する形で創薬ターゲット分子の探索研究が含まれることが多い。

・薬理活性化化合物を探索する技術（ヴァーチャルスクリーニング技術）

数百～数万からなる化合物ライブラリからアッセイ実験を通して候補化合物を探索する代わりに、計算機上で数千～数百万の化合物ライブラリから活性化化合物を選出する技術であり、計算機の進歩に伴って2000年頃より創薬現場でも実用されるようになった。現在、ヴァーチャルスクリーニングのソフトウェアは、有償、無償のものを含め非常に多くのものが存在し、創薬現場でも特別な技術ではなくなっている。

ヴァーチャルスクリーニングは技術的な観点から、リガンドベースの方法と立体構造ベースの方法の2系統に大きく分類される。既知のリガンドや活性化化合物の情報を有する場合はリガンドベースの方法が適用可能であり、標的タンパク質の立体構造情報を有する場合は立体構造ベースの方法が適用できる。

リガンドベースの方法は、1990年代から欧米の大学研究機関において化合物の記述方法や類似化合物の探索方法の開発が行われ、それらを実装した様々な有償・無償のソフトウェアが創薬現場に提供されている。様々な手法が提案されているが、Ligand based drug design (LBDD) の枠組みで定式化できる。ある標的タンパク質（もしくは複数の標的タンパク質）の既知リガンドや活性化化合物の化学構造から活性の有無（もしくは活性の強さ）を学習した分類器（もしくは回帰式）を機械学習で構築し、新しい化合物の活性を予測する方法である。例えば、化合物の化学構造の記述子（フィンガープリントなど）や化学構造のグラフ構造を機械学習による予測モデルの入力情報とする。立体構造が未知のタンパク質にも適用可能であるが、予測精度は学習データ内の既知リガンドや活性化化合物の数に大きく依存し、数が少ないと精度が悪くなる。

立体構造ベースの方法は、タンパク質と薬剤候補化合物の結合自由エネルギーを、分子動力学 (Molecular Dynamics (MD)) シミュレーションを用いて予測する、いわゆるドッキング計算ソフトと呼ばれるものであり、欧米の大学研究機関の成果物をベースに民間会社がグラフィックユーザーインターフェース (GUI) を統合したパッケージを開発し、製薬会社に有償で広く提供されている。代表的なソフトウェアとしてはCHARMM (Chemistry at HARvard Macromolecular Mechanics)、AMBER (Assisted Model Building with Energy Refinement)、NAMD (NANoscale Molecular Dynamics program)、GROMOS (GROningen MOlecular Simulation)、GROMACS (GROningen MACHine for Chemical Simulations)、DESMONDなどが挙げられる。GROMOSとGROMACSは欧州で開発されているが、それ以外は全て米国の大学や研究機関で開発されている。様々なアルゴリズムが提案されているが、Structure-based drug design (SBDD) の枠組みで定式化できる。リガンド情報が未知のタンパク質に対しても解析可能であるが、立体構造が分かっているタンパク質にしか適用できないという弱点もある。

日本では過去にNEDOプロジェクト「医用化合物スクリーニング支援システム」(2000～2004年)において、医薬分子設計研究所 板井らにより国産ソフトの開発が試みられた。その後、NEDOプロジェクト「創薬加速に向けたタンパク質構造解析基盤技術開発プロジェクト」(2008～2012年)において、大阪大学 中

2.1

村と産総研 福西らによって、新たな国産のヴァーチャルスクリーニングソフト (myPresto) が開発された。myPresto 開発は、2013 年からは経産省管轄プロジェクト「個別化医療に向けた次世代医薬品創出基盤技術開発」に受け継がれ、現在も拡張開発を行っている。myPrestoは無償で提供されており、機能面においても海外有償ソフトと遜色ない機能を有しているが、GUIを提供していないことや競合市販ソフトが低価格になったことから、製薬会社でのユーザー数は伸び悩んでいる。スーパーコンピュータ「京」(2007年～)、ポスト「京」(「富岳」)(2014年～)を用いた戦略分野の一つに創薬応用を掲げており、「京」の利用に向けて様々なソフトウェア (GENESIS、Modylas など) が開発され、利用が始まっている。

インシリコ創薬の課題のひとつとして、標的タンパク質構造情報の限界が挙げられる。タンパク質の3次元構造解析には、長年 X 線結晶構造解析と NMR (核磁気共鳴) 法が用いられてきたが、生体内の構造を完全に再現できていないなどの課題があった。近年、クライオ電子顕微鏡を用いた原子レベルのタンパク質構造解析技術が実現してきたことから、従来の課題であった生体内の環境を反映した精緻なタンパク質構造を得ることができるようになった。特に、従来結晶を得ることが難しかった膜タンパク質、タンパク質複合体などの構造解析、また生体内のダイナミックな構造変化などをとらえることが可能になったことは、インシリコ創薬の発展に大きく貢献すると期待される。AMEDが2017年度に開始した「創薬等先端技術支援基盤プラットフォーム (BINDS)」では、放射光施設、クライオ電子顕微鏡などを整備・維持し、構造解析、タンパク質生産などに高度な技術を持つ最先端研究者のサポートも提供している。製薬産業が活用することも可能であり創薬研究の加速に寄与するものと期待される。

・ 活性化化合物を最適化する技術 (リード最適化技術)

上記のスクリーニングでヒットした化合物の化学構造を合成展開し、標的タンパク質にのみ強く結合するリード化合物を合理的にデザインすることを目的として、タンパク質と化合物との結合親和性を推定する技術が開発されている。結合親和性を推定する技術も、ヴァーチャルスクリーニングと同様、リガンドベースの方法と立体構造ベースの方法の2系統に大きく分類され、さまざまな計算法が考案されている。

リガンドベースの結合親和性予測法としては、定量的構造活性相関 (QSAR) をベースとする様々な方法が国内外で開発されており、主に欧米で開発されたソフトウェア (有償・無償) が創薬現場で用いられている。一方、立体構造ベースの結合親和性予測法は、量子化学的手法と分子動力学計算的手法の2つのアプローチで開発されている。量子化学的手法では、米国カープラスら (当時ハーバード大学) が開発したQM/MM (Quantum Mechanics/Molecular Mechanics) 法 (2013年ノーベル化学賞) や我が国の北浦ら (現神戸大学) が開発したフラグメント分子軌道 (Fragment Molecular Orbital (FMO)) 法などがあり、分子動力学計算的手法では、藤谷ら (現 東京大学) が考案したMP-CAFEE (Massively Parallel Computation of Absolute binding Free Energy with well-Equilibrated states) 法や米国コールマンら (当時カリフォルニア大学) が開発したMM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) 法などがある。これらの計算法は、民間会社や大学研究機関などによってソフトウェア化 (有償、無償ともにあり) されている。わが国では、MP-CAFEE 法がスーパーコンピュータ「京」における戦略ソフトとして開発された経緯がある。

・ ADMET を予測する技術

薬物の体内における吸収 (Absorption)、分布 (Distribution)、代謝 (Metabolism)、排泄 (Excretion)、毒性 (Toxicity) を予測する技術であり、前臨床や臨床フェーズでの薬物の有効性・副作用予測や薬物投与

量の推定に重要な技術である。ADMET 予測は創薬過程で候補化合物のデザインに用いられている技術であるが、薬物動態の予測 (シミュレーション) は前臨床・臨床段階における薬物投与計画に用いられることがほとんどで、薬物のドラッグデザインに直接フィードバックできる計算技術には至っていない。いずれの方法も国内外の大学研究機関で予測モデルの開発がなされ、無償での公開や民間会社から製品サービスが行われている。

・既存薬の新しい適応可能疾患を予測するドラッグリポジショニング技術

疾患感受性遺伝子の情報を、遺伝統計解析を通じて多彩な生物学・医学データベースと分野横断的に統合することにより、新たな疾患病態や組織特異性の解明、疾患バイオマーカーの同定、疾患疫学上の謎の解明などに貢献できることが明らかとなりつつある。ファイザー製薬が資金提供をし、米国 BROAD Institute が上市薬を中心とした 1,309 種類の低分子化合物に対する遺伝子群の発現パターン (signature) 変動解析を実施し、遺伝子発現データベースおよび解析用ソフトウェア Connectivity Map (C-Map) として公開した¹⁾。この結果は BROAD Institute の Web 上でフリーにアクセス可能である。用いた細胞種は数種類と限られているが、上市薬を統一的方法で網羅的にリプロファイリングしたという意味では先駆的な研究である。自身の興味ある化合物の分子 signature を実験で取得し、C-Map 上のデータと比較することにより、類似した signature を有する上市薬を見つけることが可能である。その結果、その上市薬に付加している情報からリポジショニング、あるいは薬効パスウェイと副作用予測などの情報を得ることが出来る。例えば、炎症性大腸炎、前立腺がん、大腸がんに対して潜在的な効能がある既存薬や医薬品候補化合物の発見が多く報告されている^{2), 3)}。実際に日本においても、C-Map を活用し治療抵抗性の前立腺ガンに有効な化合物を C-Map 上から見だし、医師主導型の臨床研究としてスタートしている。

解析している薬物の処理前後のオミクスデータを活用し、パスウェイを推定・絞り込みが可能である。これまで最も利用されているオミクスデータは発現頻度解析 (トランスクリプトーム) であるが、質量分析によるタンパク質発現量絶対定量解析、あるいはリン酸化定量解析などの翻訳後修飾の網羅的な解析から、低分子化合物の生物活性を説明するパスウェイを効率的かつ体系的に抽出可能となってきた。日本でもトキシコゲノミクス・インフォマティクスプロジェクトにより、Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs) が開発され、約 150 個の薬物のラット個体およびラット・ヒト肝細胞への曝露時の遺伝子発現データが保存されている⁴⁾。

しかしながら、C-Map や TG-GATEs は化合物や細胞種の数に限られていた。C-Map における細胞種は主に 3 種類だけであった。近年、米国 BROAD Institute において、C-Map の後継プロジェクトとして Library of Integrated Network-based Cellular Signatures (LINCS) データベースが開発された⁵⁾。LINCS では、約 2 万個の薬物 (既承認薬以外の化合物が大半) と 77 種類のヒト細胞株に対する遺伝子発現データが保存されている。化合物添加に対する遺伝子発現プロファイルだけでなく、遺伝子のノックダウンまたは過剰発現に対する応答を表す遺伝子発現プロファイルなども利用可能である。これによって、多様な細胞種における膨大な数の化合物のトランスクリプトーム情報を解析できるようになり、このリソースの創薬応用が大いに期待されている。さらに最近データが更新され、32,855 個の化合物を 99 個のヒト由来細胞株に対する遺伝子発現データに拡張されている (2021 年 3 月時点)。化合物が標的タンパク質を阻害するか活性化するかによって薬理効果は全く異なるものになる。そこで、化合物応答遺伝子発現プロファイルと遺伝子ノックダウン応答遺伝子発現プロファイルと比較、化合物応答遺伝子発現プロファイルと遺伝子過剰発現応答遺伝子発現プロファイルと比較して、化合物の標的タンパク質を阻害と活性化を区別して予測する手法なども提案されてい

2.1

俯瞰区分と研究開発領域
健康・医療

る⁶⁾。しかしながら、同じ薬物の遺伝子発現プロファイルであっても、細胞種、濃度、測定時点、プラットフォームの違いによって変動する点には注意する必要がある。また細胞株によってはデータがほとんど取られていないものもあり、それらの欠損データの補完が課題である。

ドラッグリポジショニングは、薬物と疾患の効能という意味での関係性の有無を予測する問題として定式化できるが、機械学習における教師付き学習の枠組みで、薬物・疾患ネットワークを予測する手法が開発されている。薬物の特徴プロファイル（薬の副作用や作用点など）と疾患の分子的機序プロファイル（病因遺伝子、環境因子、疾患パスウェイなど）を融合解析し、薬物・疾患ペアを「効能あり」のクラスか、「効能なし」のクラスに分類する機械学習アルゴリズムが開発されている^{7), 8)}。教師無し学習の枠組みでは、薬物応答の遺伝子発現プロファイルと疾患患者の遺伝子発現プロファイルを比較し、逆相関の組み合わせを見つけることで、疾患病態特異的な遺伝子発現パターンを打ち消すような薬物を選ぶ手法が提案されており、そのための様々なスコアリング手法が提案されている。疾患によっては、疾患と薬の相関は必ずしも逆相関になるとは限らないため、正の相関まで含めたスコアリングの手法なども提案されている⁹⁾。

薬の分子は本来目標とした標的タンパク質だけではなく、それ以外の複数のタンパク質（オフターゲット）に結合することで予想外の薬理作用を起こすものがあり、期待する薬効だけではなく様々な副作用を引き起こす。単一の標的タンパク質だけでなく全ての標的タンパク質（プロテオーム全体）を考慮するポリファーマコロジー（polypharmacology）の概念に基づき、polypharmacologyプロファイルを入力とし、疾患情報と関連付けることによって、薬物の適応可能疾患を機械学習で予測する手法が提案されている¹⁰⁾。上記の手順を行うためには、薬物の潜在的な標的分子をオフターゲットも含めて全て推定する必要があるが、これまでに薬物の標的分子をゲノムワイドに予測するための機械学習の手法が多く提案されている。薬物の入力データとしては、薬物化学構造^{11), 12)}、薬物副作用^{13), 14)}、市販後薬害事象¹⁵⁾、薬物応答遺伝子発現プロファイル^{16), 17)}などが利用されている。予測モデルの学習アルゴリズムとして、ロジスティック回帰、最近隣法、距離学習、カーネル法、ディープラーニングなどが提案されている。また複数の薬の組み合わせによって効能を増強させる研究や、逆に複数の薬の組み合わせによる重篤副作用を予測する研究なども近年行われてきており、機械学習の技術も利用されている。

複雑系薬剤の典型例である漢方薬や生薬の作用機序をポリファーマコロジーの観点から情報科学的に解剖する研究や、漢方薬や生薬のリポジショニングの研究も近年報告され始めている。漢方医薬ビッグデータを解析し、漢方薬が生体に薬理学的効果を及ぼす作用機序の考察を可能にするデータベースKampoDBが富山大学和漢医薬学総合研究所のグループにより開発されている¹⁸⁾。漢方薬の主要成分化合物に関して、ヒトのタンパク質立体構造に対するin silico結合シミュレーションを実施し、その結果（結合エネルギー、結合状態の分子イメージなどの物性値情報）を分かりやすく表示する機能をデータベース上に構築している。数百万個の化合物・タンパク質間相互作用ペア情報を用いた教師付き学習アルゴリズムと化学構造類似性検索を組み合わせることで、漢方薬の成分化合物の標的タンパク質候補を探索した結果もまとめられている。様々な疾患に対する治療標的の情報を基に、漢方薬の新しい適応可能疾患を予測し、これらの解析結果を視覚化するシステムとなっている。腹痛や腸管運動機能異常に効果がある漢方薬「大建中湯」の抗炎症作用や炎症関連大腸発癌に対する抗癌作用を予測し、その作用機序を複数のパスウェイの複合的な制御である可能性を示唆するなどの事例が報告されている。

(4) 注目動向

[新展開・技術トピックス]

・MDシミュレーション専用計算機

米国では、実業家 D. E. Shaw が自己資金で分子動力学計算専用のスーパーコンピューター開発を行い、これに Microsoft 社ビル・ゲイツやメガファーマの支援も受けて、2008 年にスパコン ANTON を発表した。現在は ANTON 2 として、数十万原子を扱うことが可能でミリ秒スケールの生体高分子の分子動力学計算を実現している。わが国では、理研 泰地らが MDシミュレーション専用計算機 MDGRAPE-4A を開発している。SARS-CoV-2 のメインプロテアーゼと既存のヒト免疫不全ウイルスプロテアーゼ阻害薬の結合を 6 μs まで観察し、薬分子と標的タンパク質の結合が柔軟であることなどを報告している¹⁹⁾。

・テンソル分解アルゴリズムを用いた予測

疾患や医薬品に関するオミクスデータの構造が複雑化し、行列形式のデータだけでなく、テンソル構造のデータが創生されている。シングルセルレベルでのオミクスデータも得られるようになり、データ構造が複雑化している。これまでの統計手法や機械学習手法は行列データが主に対象であったが、テンソル構造のデータを扱うための機械学習手法が提案されている。例えば、テンソル分解を用いた次元削減や欠損値・未観測値の補間法などが提案されている。補完した薬物応答遺伝子発現データの解析によって、治療薬探索の予測精度を向上できることが報告されている²⁰⁾。

・データマイニング技術によるドラッグリポジショニング、併用療法探索

臨床投薬データから、特定の疾患と投薬頻度をもとにオッズ比を計算し、オッズ比が低い薬を疾患予防薬の候補として推定するデータマイニング技術が提案されている^{21), 22)}。またその技術を応用し、特定の薬の毒性を緩和する薬の組み合わせの提案も可能である。

・分子間相互作用ネットワークに基づく標的分子、医薬品探索

疾患の病態は基本的に疾患特異的と考えられるが、特徴的な分子機序は疾患間でも共通し、創薬標的やバイオマーカーが共通する場合がある。それゆえ、疾患間の共通性の理解は、医薬品の探索や最適な治療法の選択に直結する。例えば相関がある疾患として、癌とリウマチ、糖尿病とアルツハイマー病、喘息とアトピー性皮膚炎などの関係が知られている。疾患感受性遺伝子と疾患治療薬の標的遺伝子の分子ネットワーク上(タンパク質間相互作用や代謝パスウェイ)での近さやトポロジーによる疾患関係性の研究や、疾患患者の遺伝子発現プロファイルの類似性による疾患関係性の研究、電子カルテデータによる疾患関係性の研究などが報告されている。感受性遺伝子群や発現異常遺伝子群から疾患を特徴付ける分子ネットワーク(代謝パスウェイ、シグナル伝達、タンパク質間相互作用、遺伝子制御ネットワークなど)を疾患間で比較し、疾患間の分子機序の類似性をシステムレベルで評価することによって、疾患間の共通性や特異性をより正確に見出す手法が提案されている。さらに分子ネットワークに基づく疾患類似性を応用し、情報が少ない疾患に対する創薬標的分子や治療薬を探索する機械学習アルゴリズムなども報告されている²³⁾。

・分子ネットワーク解析によるドラッグリポジショニング、併用療法探索

複数の薬剤の組み合わせによる相乗効果を活用した化学療法が、がんや神経変性疾患など多因子疾患に対する有効な治療法として注目されている²⁴⁾。最近、分子ネットワークを用いた薬物の組み合わせ予測手法が

提案されている²⁵⁾。感受性遺伝子情報を基に疾患を特徴付ける分子ネットワークを同定し、複数の薬の標的分子群の分子ネットワーク上での生体分子群の重なり程度をトポロジーの観点から評価することで薬物の最適な組み合わせを予測する教師なし学習アルゴリズムである。疾患の分子ネットワークと薬物群の分子ネットワークの重なりタイプによって、相乗効果や副作用など様々な作用に結びつける。ただし、論文中では高血圧症でのみの有効性が確認されているだけであり、他の疾患に対する有効性は定かではない。教師付き学習アルゴリズムとして、相乗効果の知られている薬物ペアに頻出する分子的特徴を抽出するアルゴリズム、複数の薬物の標的分子プロフィールや疾患プロフィールの組み合わせをスパースモデリングで融合する機械学習法などが提案されている²⁶⁾。薬物の化学構造を内部グラフ、薬物ペアのネットワークを外部グラフとみなし、グラフ畳み込みニューラルネットワークを二重にモデル化する Graph of graph (GoG) ニューラルネットワークなども提案され、薬物間相互作用や薬物間相乗効果を予測する手法なども提案されている²⁷⁾。

・パスウェイ創薬

生体分子間相互作用ネットワークやパスウェイの情報を医薬品候補化合物の探索に利用する動きがある。従来の創薬では、1つの生体分子を制御する薬を探索するアプローチが一般的だが、生体分子間の相互作用の影響を考慮できないので、1つの生体分子を制御が期待する薬効に必ずしも繋がらないという問題がある。生体分子相互作用ネットワークの機能モジュールであるパスウェイを制御する薬を探索する方法が提案されている²⁸⁾。大規模な薬物応答遺伝子発現データと生体分子相互作用ネットワークの融合解析により、薬物に回答してがん特異的に活性化もしくは不活性化されるパスウェイを推定し、がん関連パスウェイを制御する化合物を予測する手順になる。ある統合失調症治療薬が、期待する抗がん作用をもつことが実験的にも確認でき、インシリコ予測の妥当性を示されている。

・先制医療への展開

富山大学の研究グループを中心に、疾病の状態遷移の臨界状態である「発症前の未病状態や再燃前状態」での医療介入という先制医療の研究を進めている。疾病が発症してから治すという従来の治療戦略とは根本的に異なる新規治療戦略であり、疾病が発症しないように予防するという予防医療の観点からも注目される。そのために、オミクスデータや分子ネットワークなどの生体情報から未病・再燃前状態を数理科学的に客観的に予測する解析手法を構築するとともに、疾病形質が既に獲得された状態での薬剤介入ではなく、未病・再燃前状態での薬剤介入のための新たな未病創薬コンセプトの理論構築とその実証を目指している。未病状態を制御するツールとして日本の伝統的な薬である漢方薬の有効活用している。ネットワークバイオマーカーなどの数理的アプローチは東京大学の研究グループとの共同研究によって理論構築を進めている²⁹⁾。

・AI創薬

ディープラーニング（多層ニューラルネットワークによる深層学習）や強化学習（将来的な利益を最大化するように試行錯誤を通じて行動を決定する学習法）などAI基盤技術の進展に伴い、これらを創薬に応用しようという試みがある（AI創薬）。これまでも、サポートベクターマシンに代表されるカーネル法、lassoなどのスパースモデリング、ランダムフォレスト、ディープラーニングなどの機械学習の手法は既に創薬工程の様々な問題（化合物の活性予測、化合物の標的予測、化合物の薬理作用・副作用予測など）で提案されてきている。Merck Molecular Activity Challengeと呼ばれる世界的な化合物活性予測コンテストで、様々な機械学習の手法が比較および検討された。その結果、1位はディープラーニングだったが、2位のサポートベク

ターマシシとほとんど差は無かった。化合物活性予測の精度に関するディープラーニングの効果は限定的と言わざるをえない。ディープラーニングは常に万能ではなく、特徴データの形式、学習データの質・量、解くべき問題に大きく依存すると思われる。生命医学や創薬科学の関連データは欠損値が大部分で、バイアスが大きく、サンプル数も少ないので、そのあたりの独特の難しさが障壁となる。

ディープラーニングの化合物活性予測への応用では、画像認識で有効性が知られている畳み込みニューラルネットワーク Convolutional neural network (CNN) を化合物のグラフ構造に拡張した、グラフ畳み込みニューラルネットワーク Graph convolutional neural network (GCNN) の研究開発が進んでいる³⁰⁾。単一の標的分子だけでなく複数の標的分子に対する化合物の活性を予測するマルチタスク学習の枠組みなども応用されている。化合物データを解析する深層学習モデルを実装した DeepChem などのライブラリなどが開発されたこともあり、創薬現場でも利用する人が増えている³¹⁾。化合物の化学構造記述子やフィンガープリントなど人間が設計する特徴に依存せず、GCNN は必要な特徴を化合物の化学構造を表すグラフから自動的に構築できることが利点とされているが、これまでもグラフ構造を入力とする機械学習手法 (グラフカーネルによるサポートベクターマシンなど) はあるので厳密な性能の比較が望まれる³²⁾。

グラフ畳み込みニューラルネットワークは、創薬だけではなく、健康食品開発にも応用されている³³⁾。食物成分は栄養素としてだけでなく人体の生理学的機能のモジュレーターとしても機能し、科学的にも産業的にも注目の最前線にある。例えば、食物成分には、抗菌作用や血圧抑制効果、免疫調整効果などがあるものがある。しかし、生物活性を示す食品成分のほとんどは作用機序が解明されておらず、潜在的な健康効果が眠っている可能性がある。食品成分化合物の潜在的な標的分子群を推定し、健康効果の予測に繋げる機械学習手法が報告されている。

目標の特性 (薬効など) を持つ化合物を、大量の化合物のスクリーニングから見つけるのではなく、目標の特性を持つ化学構造を新しく予測する逆構造活性相関解析 (通常の構造活性相関解析とは逆方向のアプローチ) も最近研究が進められている。化合物の構造を Simplified Molecular Input Line Entry System (SMILES) という文字列で表記し、確率的言語モデルを用いて文字列のパターンを学習する研究が、創薬やマテリアルインフォマティクスの分野で出現している。深層学習を応用したニューラルネットワークモデルによる医薬品候補化合物の設計手法が、近年特に注目されている³⁴⁾。SMILES に基づく構造生成の先行研究事例として、N-gram という言語モデルや変分自己符号器や再帰的ニューラルネットワークなどディープラーニングによる言語認識・生成を用いた分子設計手法、モンテカルロ木探索と再帰的ニューラルネットワークを組み合わせた構造発生手法 ChemTS など提案されている。他にも深層学習を応用した構造生成の先行研究として、変分オートエンコーダ³⁵⁾ や敵対的生成モデルなどを応用した Chemical VAE (Variational Autoencoder) や Grammer VAE、DruGAN などの手法が報告されている^{36), 37), 38), 39)}。遺伝子発現プロファイルの情報を使う化合物生成モデルなども提案されている⁴⁰⁾。合成可能性を考慮し、複数の化合物から一つの化合物を生成するための手法 Molecular Chef など提案されている⁴¹⁾。しかしながら、提案される化学構造が実際に合成可能なものかどうかは保証が無いのが現状である。

化合物の構造最適化は創薬における重要課題の一つである。最近、活性化合物の最適化に繋がる SAR Matrix と呼ばれる情報技術がドイツのボン大学の Jürgen Bajorath によって提案されている⁴²⁾。化学構造の部分構造変換パターンと生物活性情報の対応を地図として視覚化するようなソフトウェアも日本の民間企業である理論創薬研究所によって開発されている⁴³⁾。

[注目すべき国内外のプロジェクト]

- ・ Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) Consortium

マサチューセッツ工科大学が中心となり、2018年に立ち上がった医薬品発見と合成のための機械学習コンソーシアムである。Pfizer社、Novartis社、Eli Lilly社など大手製薬会社も参画している。同コンソーシアムは、低分子の発見と合成に役立つソフトウェアのデザイン促進を目的としている。

- ・ Machine Learning Ledger Orchestration for Drug Discovery (MELLODDY) project

欧州を中心とした製薬企業10社とアカデミア、IT企業が参画して2019年から開始された。機械学習による有望な化合物予測プラットフォームの構築を目指している。各企業はFederated Learningという学習手法を用いることで、データの秘匿性を保持しながら1000万を超える低分子化合物のデータを共同利用できる枠組みを構築し、世界最大のコレクションを活用した、より正確な予測モデルの確立、創薬の効率化が期待されている。

- ・ LINC (Life Intelligence Consortium)

わが国では、2016年に製薬・化学・食品・医療・ヘルスケア関連のライフサイエンス分野のためのAI・ビッグデータ技術開発をpre-competitive領域で開発し、産業競争力を高めるコンソーシアムとしてLINCが京都大学奥野を代表として設立された。アカデミア、製薬・ライフサイエンス企業、IT企業など100以上の機関が参画し、10組のワーキンググループがシームレスなAI創薬プラットフォームの構築を目指した技術開発を進めた。成果は医薬基盤・健康・栄養研究所内のサーバーにまとめられ、LINCメンバーに共有されている(LINC DB)ほか、文献情報を基にした研究者探索サービスJDream Expert Finderが製品化された。2021年にはポストLINCを立ち上げ、競争領域への移行を目指した活動が継続される予定である。

- ・ 官民研究開発投資拡大プログラム (PRISM) 「新薬創出を加速する人工知能の開発」

「創薬標的の枯渇」問題を克服するための取り組みとして、医薬基盤・健康・栄養研究所、理化学研究所、科学技術振興機構など17の産学官研究機関が参画して2018年より開始されている。特発性肺線維症と肺がんにおける新規創薬ターゲット探索を目的に、臨床情報と文献など既存知識のデータベースを構築し、それらを活用して創薬ターゲットを推定するAIの研究開発が行われている。

- ・ AMED創薬支援推進事業「産学連携による次世代創薬AI開発」 -

2020年度から開始された深層学習を始めここ数年間のめざましいAI技術の発展を取り込み、化合物の最適化を飛躍的に効率化する創薬AIのためのソフトウェアを開発するプロジェクトである。複数の生体分子に対する化合物の活性予測AIと構造発生AIの融合が特色である。製薬企業に蓄積した化合物に関する生物活性データや毒性データを最大限に活用し、産学双方での創薬技術の向上を図る予定である。

- ・ AI-SHIPS (AI-based Substances Hazardous Integrated Prediction System)

経済産業省で、毒性関連ビッグデータを用いた人工知能による次世代型安全性予測を目指したAI-SHIPS(プロジェクト正式名称:省エネ型電子デバイス材料の評価技術の開発事業(機能性材料の社会実装を支える高速・高効率な安全性評価技術の開発・毒性関連ビッグデータを用いた人工知能による次世代型安全性予測手法の開発))プロジェクトが実施されている。基本的には一般化合物の毒性が予測対象となっており、動

態的アプローチ、代謝的アプローチ、AI的アプローチ、トキシコゲノミクスのアプローチを組み合わせ、統合的予測システムの構築を目指している。同じ手法は医薬品開発における毒性研究にも利用可能なので、創薬応用の視点からも研究成果が期待されている。

(5) 科学技術的課題

・ 医薬品の標的となる疾患原因分子（主にタンパク質）の探索技術

患者検体などの疾患サンプルとそれらを測定したオミクスデータが必要であり、実験データの収集体制を前提とした上で、技術開発をしなければならない。患者検体の数は限られるため、画像データ解析のような膨大なサンプル数に基づく情報技術ではなく、少数のサンプル数に基づく情報技術の開発が必要不可欠である。

ゲノム、トランスクリプトームレベルでの疾患研究は多く存在するが、プロテオームレベルの疾患研究は立ち遅れている。創薬応用まで目指す場合は薬物の直接ターゲットとなるプロテオームレベルの情報収集が必須となる。しかしながら、プロテオミクス技術で測定できるタンパク質の数は限られているため、トランスクリプトームレベルの網羅性には到底及ばない。本来最良の創薬標的タンパク質がプロテオームデータに入らず見逃される可能性がある。

ゲノム、トランスクリプトーム、プロテオームだけでなく、エピゲノム、メタボローム、臨床情報など多階層のオミクス情報の連結による、創薬標的分子の探索が期待されている。しかしながら、同一患者に対して多階層のオミクス情報や臨床情報を測定することはコストが非常に大きくなる。多階層のオミクス情報を連結する情報技術は確立されておらず、その開発は急務である。また各階層のオミクス情報は欠損値や未観測値が非常に多く、そのデータ補完は大きな課題である。

・ 薬理活性化化合物を探索する技術（ヴァーチャルスクリーニング技術）

現状の計算技術の予測精度は平均で5%～10%程度であり、予測精度向上のための方法論の更なる改良が必要である。

リガンドベースの方法では、病態情報、標的タンパク質情報、薬理活性情報、副作用情報などを総合的に考慮する技術開発の可能性がある。特に、ビッグデータ解析技術の応用が期待できる。

医薬品候補化合物取得の方法に予め標的を定めないフェノタイプベースのハイスループットスクリーニング（PHTS）があるが、目的とするフェノタイプに合致した化合物をヒット化合物として得るための重要な方法である。一方で、PHTSで見出された化合物は標的分子や作用機序が未知で、標的分子の同定に大きな労力を必要とするため、標的未知化合物の標的予測法が切望されている。化合物の化学構造、人体に対する副作用パターン、化合物応答遺伝子発現情報などを基に機械学習の技術を駆使した標的分子予測法が提案されてきている。しかしながら、リガンド情報が未知なタンパク質に対する予測（学習データが無い場合の予測）は困難なのが現状である。

・ 活性化化合物を最適化する計算技術（リード最適化技術）

化合物とタンパク質との結合親和性予測において、さまざまな標的分子（タンパク質以外も含む）に対して、高精度で予測する性能を担保するために、方法論の更なる改良が必要である。特に、現在の計算法では計算時間がかかり過ぎるため、計算機コストなどの面においても創薬現場で容易に利用できる技術ではない。

・ ADMET を予測する計算技術

現状の ADMET 予測技術は、予測精度も低く、信頼性が低い。また、新規化合物への適用性において疑問がある。副作用の予測では、臨床現場での副作用情報を加味した予測技術はない。薬物動態の予測（シミュレーション）は前臨床・臨床段階における薬物投与計画に用いられることがほとんどであり、分子デザイン段階での応用研究はない。モデル動物を用いた薬効薬理試験は臨床予測性が不十分であり、医薬品の研究開発の成功率が向上しない大きな原因の一つになっている。

・ 既存薬の新しい適応可能疾患を予測するドラッグリポジショニング技術

疾患によって得られるゲノムデータやオミクスデータ、臨床データの質や量の差が大きく異なる。患者数が多い疾患はデータが豊富に得られるのに対し、希少疾患に対するデータは非常に少ない。同様に、薬に関しても、利用できるオミクスデータの質や量の差が大きく異なる。薬物投与時の臨床情報は有用なリソースとなるが、日本では大規模な臨床データが手に入りにくい。

・ 機械学習による医薬品候補化合物の構造生成

深層学習による医薬品候補化合物の設計手法が、近年特に注目されている。変分オートエンコーダや敵対的生成モデルなど様々な手法が使われているが、いずれの手法も化学構造の変換が不確実であり、生成する構造の多様性や、生成した分子の生物活性の強さ、毒性の考慮などがされておらず、実用性には多くの課題を有している。しかしながら、提案される化学構造が実際に合成可能なものかどうかは保証が無いのが現状である。合成可能性の指標として、SA (Synthetic Accessibility) スコアという指標がよく利用されているが、改善の余地は大きい。

(6) その他の課題

- ・ オミクス研究はデータ測定におけるコストが高くなる。設備の共用などプラットフォーム作りが求められる。分子リプロファイリングの基盤を構築する解析機器・計算機リソースなどは一企業が単独保有運営するには負担が大きく、pre-competitive に全製薬企業が利用可能な拠点を構築する政策が必要である。
- ・ データの質を担保することが非常に重要である。健康人のベースラインデータや生活習慣との関連付け、サンプル間を比較するための内部標準物質の設定や絶対定量系の確立など、超えなければならない壁は多い。海外の競合相手も同様の問題を抱えているため、産官学で連携した対策とそれに即した実行が必要である。
- ・ データの標準フォーマットやオントロジーの整備が課題である。遺伝子名、化合物名、疾患名、症状、フェノタイプ、副作用の定義、データフォーマットなどがデータベース間や研究者コミュニティによって異なるため、多様なリソースのデータを統合解析する際には、情報解析可能にするためのデータクリーニングに膨大な時間を費やす。
- ・ 創薬インフォマティクス研究は、医学薬学の高度な知識と計算科学の専門性が要求されるため、専門人材が産業界で枯渇している。また、学際領域の教育システムが定着しにくいわが国においては、総合的な創薬の教育体制を整備している大学は皆無にひとしい。そもそも日本の大学において創薬インフォマティクスを専門にする教員のポストはほとんど無い。大学で創薬研究をした学生や研究員は、企業に流れる傾向がある。

(7) 国際比較

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	○	→	<ul style="list-style-type: none"> ・ 前述のPRISMや2018年に設立されたがんゲノム情報管理センターなど、各種オミックスデータと臨床データをリンクして収集、活用する仕組みが構築されつつある。 ・ 製薬企業が非競争領域で化合物データなどを共有、AMED事業LINCなどでAIシステム構築を促進している。
	応用研究・開発	○	→	<ul style="list-style-type: none"> ・ IT企業の創薬分野への進出、AI創薬ベンチャーが増加し、製薬企業との協働が進んでいる。 ・ LINCが競争領域での実装に向けて活動を継続する。
米国	基礎研究	◎	→	<ul style="list-style-type: none"> ・ 2017年にGlaxoSmithKline社はエネルギー省傘下のLawrence Livermore National Laboratory、国立がん研究所傘下のFrederick National Laboratory等とAccelerating Therapeutics for Opportunities in Medicine (ATOM) コンソーシアムを立ち上げた。最新のITテクノロジー、多様な生物学的データなどを用いて創薬を劇的に効率化するプラットフォームの構築を目指している。
	応用研究・開発	◎	↗	<ul style="list-style-type: none"> ・ アカデミアの成果などを基にAtomwise社、InveniAI社、twoXAR社など本分野でベンチャーが数多く立ち上がっている。 ・ Stanford大Pande研の成果をベースに2019年Genesis Therapeutics社が設立、グラフ畳み込みニューラルネットワークアルゴリズムPotentialNetによる標的-リガンド親和性、ADMET予測などが注目され、2020年にはGenentech社と提携した。
欧州	基礎研究	◎	→	<ul style="list-style-type: none"> ・ 英国国民保健サービス (NHS) はGoogle社傘下のDeepMind社と提携し、医療分野でのAI活用を推進している。 ・ 2019年にNHSのデジタル戦略責任組織NHSX傘下に2億5,000万ポンドを投じてAI Labを設立し、ヘルスケア領域でのAI開発をリードすることを目指している。
	応用研究・開発	◎	↗	<ul style="list-style-type: none"> ・ 英Exscientia社はAI創薬プラットフォームCentaur Chemistを活用し、大日本住友製薬とDSP-1181を共同開発、AIを使ってデザインした化合物として初の臨床試験を開始した。 ・ 英BenevolentAI社は、文献情報等を独自のAIプラットフォームで解析し、関節リウマチ治療薬BaricitinibがCovid-19治療薬となる可能性を示した。
中国	基礎研究	○	↗	<ul style="list-style-type: none"> ・ 北京大学・中国科学院上海薬物研究所などが中心となり、創薬応用に向けた研究が数多く行われている⁴⁴⁾。
	応用研究・開発	◎	↗	<ul style="list-style-type: none"> ・ Baidu社、Tencent社、Huawei社、ByteDance社等の中国テックジャイアントAI創薬への参入を表明している。 ・ Insilico Medicine社、XtalPi社、neoX社などAI創薬ベンチャーが成果を報告、大手製薬企業との提携、資金調達を進めている。
韓国	基礎研究	△	→	<ul style="list-style-type: none"> ・ 韓国科学技術院は薬物と薬物、人間と薬物の相互作用を正確に予測する「DeepDDI」³⁷⁾、配列情報から酵素機能を表現するEC番号を正確に予測する「DeepEC」を開発した⁴⁵⁾。
	応用研究・開発	△	→	<ul style="list-style-type: none"> ・ SK Biopharmaceuticals社、Standigm社がAIを活用した創薬に参入している。

2.1 俯瞰区分と研究開発領域
健康・医療

(註1) フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

(註2) 現状 ※日本の現状を基準にした評価ではなく、CRDS の調査・見解による評価

◎：特に顕著な活動・成果が見えている

○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない

×：特筆すべき活動・成果が見えていない

(註3) トレンド ※ここ1～2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考・引用文献

- 1) Lamb J. et al., “The Connectivity Map : Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease”, *Science* 313, no. 5795 (2006) : 1929–1935. DOI : 10.1126/science.1132939.
- 2) Dudley, J. T. et al. *Sci. Transl. Med.* 3, 96ra76, 499, 2011.
- 3) Kosaka, T. et al., “Identification of drug candidate against prostate cancer from the aspect of somatic cell reprogramming”, *Cancer Sci.* 104 (2013) : 1017 – 1026, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7657195/> (2021年2月3日アクセス) .
- 4) Igarashi Y., et al., “Open TG-GATEs : a large-scale toxicogenomics database”, *Nucleic Acids Res.* 43 (2015) : D921-D927, <https://academic.oup.com/nar/article/43/D1/D921/2439524> (2021年2月3日アクセス) .
- 5) Subramanian A. et al., “A Next Generation Connectivity Map : L1000 Platform and the First 1,000,000 Profiles”, *Cell* 30 171, no. 6 (2017) : 1437-1452.
- 6) Sawada, R. et al, “Predicting inhibitory and activatory drug targets by chemically and genetically perturbed transcriptome signatures”, *Scientific Reports* 8 (2018) : 156.
- 7) Gottlieb, A. et al., “PREDICT : a method for inferring novel drug indications with application to personalized medicine”, *Mol. Syst. Biol.* 7 (2011) : 496.
- 8) Iwata H., et al., “Systematic Drug Repositioning for a Wide Range of Diseases with Integrative Analyses of Phenotypic and Molecular Data”, *J. Chem. Inf. Model.* 55, no. 2 (2015) : 446–459.
- 9) Cheng J et al., “Systematic evaluation of connectivity map for disease indications”, *Genome Med.* 6, no. 12 (2014) : 540. DOI : 10.1186/s13073-014-0095-1.
- 10) Sawada R., et al., “Target-Based Drug Repositioning Using Large-Scale Chemical-Protein Interactome Data”, *J. Chem. Inf. Model.* 55, no. 12 (2015) : 2717–2730.
- 11) Yamanishi, Y. et al., “Prediction of drug-target interaction networks from the integration of chemical and genomic spaces”, *Bioinformatics* 24, no. 13 (2008) : 232-240. DOI : 10.1093/bioinformatics/btn162.
- 12) Keiser M. et al., “Predicting new molecular targets for known drugs”, *Nature* 462, no. 7270 (2009) : 175-81. DOI : 10.1038/nature08506.
- 13) Campillos, M. et al., “Drug target identification using side-effect similarity”, *Science* 321, no.

- 5886 (2008) : 263-266. DOI : 10.1126/science.1158140.
- 14) Yamanishi, Y. et al., “Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework”, *Bioinformatics* 26, no. 12 (2010) : i246-i254. DOI : 10.1093/bioinformatics/btq176.
- 15) Takarabe, M. et al., “Drug target prediction using adverse event report systems : a pharmacogenomic approach”, *Bioinformatics* 28 (2012) : i611-i618. DOI : 10.1093/bioinformatics/bts413.
- 16) Hizukuri Y. et al., “Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner”, *BMC Medical Genomics* 8 (2015) : 82.
- 17) Iwata M., et al., “Elucidating the modes of action for bioactive compounds in a cell-specific manner by large-scale chemically-induced transcriptomics”, *Scientific Reports* 7 (2017) : 40164.
- 18) Sawada et al, “KampoDB, database of predicted targets and functional annotations of natural medicines”, *Scientific Reports* 8 no. 1 (2018) : 11216. DOI : 10.1038/s41598-018-29516-1.
- 19) KOMATSU, Teruhisa S. et al., "Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation", *Scientific Reports*, 10 (2020) : 16986.
- 20) Iwata, M. et al, “Predicting drug-induced transcriptome responses of a wide range of human cell lines by a novel tensor-train decomposition algorithm”, *Bioinformatics* 35 (2019) : i191–i199. DOI : 10.1093/bioinformatics/btz313.
- 21) Hashikawa N et al, “HSP105 prevents depression-like behavior by increasing hippocampal brain-derived neurotrophic factor levels in mice”, *Science Advances* 3, no. 5 (2017) : e1603014. DOI : 10.1126/sciadv.1603014.
- 22) Horinouchi Y. et al, *Scientific Reports* 8, no.1 (2018)
- 23) Iida, M. et al, “Network-based characterization of disease–disease relationships in terms of drugs and therapeutic targets”, *Bioinformatics* 36 (2020) : i516–i524. DOI : 10.1093/bioinformatics/btaa439.
- 24) Sun et al, *Sci. Transl. Med.*, 5, 205rv201, 2013
- 25) Cheng et al, “A genome-wide positioning systems network algorithm for in silico drug repurposing”, *Nature Commun* 10 (2019) : 1197.
- 26) Iwata, H. et al, “Large-Scale Prediction of Beneficial Drug Combinations Using Drug Efficacy and Target Profiles”, *J. Chem. Inf. Model.* 55, no. 12 (2015) : 2705–2716. DOI : 10.1021/acs.jcim.5b00444.
- 27) Harada et al, “Dual graph convolutional neural network for predicting chemical networks”, *BMC Bioinformatics* 21 (2020) : 94.
- 28) Iwata M., et al., “Pathway-Based Drug Repositioning for Cancers : Computational Prediction and Experimental Validation”, *Journal of Medicinal Chemistry* 61 (2018) : 9583 – 9595.

DOI : 10.1021/acs.jmedchem.8b01044.

- 29) Koizumi et al, "Identifying pre-disease signals before metabolic syndrome in mice by dynamical network biomarkers", *Sci Rep.* 9, no. 1 (2019) : 8767. DOI : 10.1038/s41598-019-45119-w.
- 30) Altae-Tran et al., "Low Data Drug Discovery with One-Shot Learning", *ACS Cent. Sci.* 7 (2017) : 283–293. DOI : 10.1021/acscentsci.6b00367.
- 31) Ramsundar B et al, "Is Multitask Deep Learning Practical for Pharma?" *J. Chem. Inf. Model* 57, no. 8 (2017) : 2068-2076. DOI : 10.1021/acs.jcim.7b00146.
- 32) Mahé P et al, *Combinatorial Chemistry & High Throughput Screening*, 12 (4) : 409-423, 2009.
- 33) Fukunaga, I. et al, "Prediction of the Health Effects of Food Peptides and Elucidation of the Mode - of - action Using Multi - task Graph Convolutional Neural Network", *Molecular Informatics*, 39, no. (1-2) (2020) : e1900134. DOI : 10.1002/minf.201900134.
- 34) Schneider, P. et al, "Drug discovery with explainable artificial intelligence", *Nature Rev. Drug Discov.* 19 (2020) : 353–364.
- 35) Hinton G. E. et al, "Reducing the Dimensionality of Data with Neural Networks", *Science* 313, no. 5786 (2006) : 504-507. DOI : 10.1126/science.1127647.
- 36) Gómez-Bombarelli et.al., "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules", *ACS Cent. Sci.* 4, no. 2 (2018) : 268-276. DOI : 10.1021/acscentsci.7b00572.
- 37) Kusner et. al., "Grammar Variational Autoencoder", <https://arxiv.org/abs/1703.01925> (2021年2月3日アクセス) .
- 38) Yang et. al., "ChemTS : an efficient python library for de novo molecular generation", *Sci. Tech. Adv. Mater* 18 (2017) : 972-976. DOI : 10.1080/14686996.2017.1401424.
- 39) Kadurin et.al., "druGAN : An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico", *Mol. Pharmaceutics* 14 (2017) : 3098-3104. DOI : 10.1021/acs.molpharmaceut.7b00346.
- 40) Méndaz-Lucio O et al : *Nat. Commun.*, 11 : 1-10
- 41) Bradshaw et al, *Advances in Neural Information Processing Systems*, in press, 2020
- 42) Gupta-Ostermann D et al, *F1000Research* 3 : 113, 2014.
- 43) Yoshimori A, et al, *ACS Omega* 4, 7061-7069, 2019.
- 44) Ling Wang "Drug discovery in China : challenges and opportunities" *National Science Review* 5.5.768- 773 (2018) doi : 10.1093/nsr/nwy085
- 45) Jae Yong Ryu, Hyun Uk Kim, Sang Yup Lee, "Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers" *Proc, Natl. Acad. Sci. USA*, Jul 2019, 116 (28) 13996-14001; DOI : 10.1073/pnas.1821905116