

戦略的創造研究推進事業 AIP 加速研究
(AIP 加速 PRISM 研究)
研究課題「創薬標的分子の確からしさを検証する
ツール物質の探索」

研究終了報告書

研究期間 2018年8月～2021年3月

研究代表者：山西 芳裕
(九州工業大学大学院情報工学研究院、教授)

§ 1 研究実施の概要

(1) 実施概要

新薬を一個開発するのに数千億円の研究開発費と10年以上の歳月を要すると云われており、新薬開発の道のりは非常に険しく遠い。また創薬ターゲットの枯渇が深刻化しており、近年の創薬の低迷の一因となっている。既存の研究手法では創薬ターゲットの発見に限界があるため、ビッグデータやAI技術の有効活用が切望されている。本研究では、特発性肺線維症および肺がんの創薬標的分子候補を推定し、その創薬標的候補の「確からしさ」の検証実験に利用可能なツール化合物を、既承認薬や生物活性化合物を含む膨大な化合物セットの中から探索する手法の研究開発を行う。疾患患者のマルチオミクス情報、臨床情報、分子ネットワーク情報などの医薬ビッグデータを大規模解析し、創薬標的分子やツール化合物を効率的に予測する機械学習アルゴリズムを開発する。さらに、特発性肺線維症および肺がんの創薬標的候補分子を制御し、抗繊維化作用や肺がん抗がん作用を有するツール化合物を発見することによって、創薬標的候補分子の妥当性を実験的に評価する。山西グループと田部井グループがインシリコ予測の研究、三森グループと小路グループがウェット実験の研究を担当し、異なる専門分野の研究者が融合研究を行うことによって、創薬標的分子やツール化合物を効率的に発見することを狙う。

ツール化合物を予測するインシリコ手法開発では、疾患オミクスデータと化合物オミクスデータの相関解析からツール化合物候補を予測する手法やパスウェイ情報を活用して高感度に活性化化合物を検出する手法を開発した。また、化合物応答トランスクリプトームなどのオミクスデータを最新のテンソル分解アルゴリズムで解析し、ツール化合物候補を予測する手法を開発した。また、簡潔データ構造を用いてフィンガープリントをコンパクトに表現することによって、機械学習の分類モデルを学習するためのアルゴリズムを開発し、化合物-タンパク質間相互作用予測問題で、空間効率を既存手法比の大幅な向上を達成した。更に大規模な化合物データの効率的なスクリーニングを可能にするため、化合物記述子の類似度検索アルゴリズムの開発と実装を行った。創薬標的分子を予測するインシリコ手法開発では、ゲノムデータを用いた創薬標的探索、臨床データを用いた創薬標的探索、深層学習による創薬標的探索を行う手法の開発に取り組んだ。更に、疾患を特徴付ける様々な分子間相互作用ネットワーク(シグナル伝達、代謝パスウェイ、タンパク質間相互作用など)を比較し、疾患の類似性から創薬標的分子を効率よく探索する機械学習アルゴリズムを開発した。

肺がんの実験プラットフォーム構築では、様々なヒト肺がん細胞株を用いて活性評価系を構築し、既存の肺がん治療薬が耐性を持つ肺がん細胞に対して、効果のあるツール化合物を見出し、抗腫瘍作用、感受性や抵抗性を評価した。並行して、機械学習手法でその標的分子の候補も推定した。特発性肺線維症の実験プラットフォームでは、線維芽細胞を単離して活性評価系を構築するとともに、ヒトの病態に近い特発性肺線維症モデルマウスを用いた評価系を構築し、抗繊維化作用を持ついくつかのツール化合物を見出した。オミクスデータや臨床データを基に機械学習手法で推定した創薬標的分子やツール化合物の妥当性を、特発性肺線維症モデルマウスを用いて実験的に評価し、インシリコ予測の結果の妥当性を示唆する結果を得ることができた。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. オミクスデータと分子ネットワーク情報からの創薬標的探索手法の開発

概要：疾患のオミクスデータや生体分子相互作用ネットワークの比較から、さまざまな病気に対して創薬標的分子を探索する機械学習手法を開発した。まず、2つの病気が分子間相互作用ネットワーク上でどれくらい似ているかを評価する方法を提案した。次に、病気間類似性と創薬標的分子の共通性の対応関係を学習し、創薬標的分子候補を探索する機械学習アルゴリズムをカーネル法の枠組みで開発した。実際に、代謝に関する分子間相互作用ネットワークを用いて、様々な病気の創薬標的分子の共通性を探索した。例えば、特発性肺線維症と2型糖尿病が創薬標的分子を共有する可能性が予測された。実際に、近年のマウスを用いた研究において、糖尿病の経口薬が特発性肺線維症の治療に効果があることが近年の文

献で報告されており、この予測結果の妥当性を確認することができた。この成果は Bioinformatics 誌で論文を発表した。(Iida, et al, Bioinformatics, 36, i516-i524, 2020.)。国際的にも高く評価されており、バイオインフォマティクスのトップ国際会議 ISMB2020 で採択された(日本から採択された2件のうち、1件が我々のグループ)。

2. 複雑なデータ構造のオミックスデータを解析する機械学習手法の開発

概要: 最新のテンソル分解手法である tensor-train 分解アルゴリズムを用いた高階遺伝子発現データ解析手法を開発し、化合物応答遺伝子発現データの細胞ごとの未観測値や欠損値の補完に応用した。既存のテンソル分解手法よりも、データ補完性能において優れていることを確認した。特にテンソルの階数が高い化合物応答遺伝子発現データで、精度が高い(誤差が低い)ことを確認した。データ補完した化合物応答遺伝子データを用いて、様々な疾患の治療薬を予測する問題に応用した結果、データ補完しない場合に比べて精度が最大2倍向上することを確認した。これらの成果は Bioinformatics 誌で論文を発表した。(Iwata, et al, Bioinformatics, 35 i191-i199, 2019.)。国際的にも高く評価されており、バイオインフォマティクスのトップ国際会議 ISMB/ECCB2019 で採択された(日本から採択されたのは我々のグループのみ)。

3. 大規模な化合物-タンパク質相互作用予測を実現するデータ圧縮アルゴリズムの開発

概要: 大規模な化合物のスクリーニングを可能にするため、機械学習の予測モデルを高速に学習するためのアルゴリズムの実装を行った。84,195,800 次元の 5,372,868 タンパク質-化合物ペアのフィンガープリントのデータから化合物-タンパク質相互作用を予測するためのモデル構築を行った。データが次元と数の側面で巨大であるために、通常の機械学習アルゴリズムを適用すると、利用メモリと学習時間の側面から適応困難である。そこで、我々は、化合物-タンパク質データを圧縮し、圧縮されたデータ上で機械学習モデルの学習アルゴリズムの開発を行った。トライ木によるデータ圧縮アルゴリズムを組み合わせ、予測モデルを学習することで、従来の手法では 131GB のメモリが必要なところを、我々の手法では 24GB のメモリで学習を行うことが可能となった。この成果は BMC Systems Biology 誌で論文を発表した。(Tabei, et al, BMC Systems Biology, 13(Suppl 2):39, 2019)。

< 科学技術イノベーションに大きく寄与する成果 >

1.

概要: 本研究の目的は、創薬標的分子やその妥当性を評価するツール化合物を効率的に発見する手法の開発であり、新薬開発を加速する科学技術イノベーションに大きく貢献する。開発手法は生体分子を制御する化合物を効率的に発見するためだけでなく、化合物が相互作用する生体分子やパスウェイを予測することが可能であるため、生物活性化合物の作用機序の解明やオフターゲットの予測など、創薬研究や生命科学における様々な課題に応用可能である。

2.

概要: AI 技術を駆使して創薬工程を効率化することによって、資源の浪費を回避、人件費の削減、開発期間の短縮、開発コストの削減に貢献できる。本研究の開発手法は特発性肺線維症と肺がんだけでなく、データが得られる様々な疾患に適用可能である。実際に多くの製薬企業と共同研究を行い、開発手法は実際の医薬品開発の工程に応用されている。

3.

概要: 本研究の対象疾患である特発性肺線維症と肺がんは致死率が高く、有効な治療法がほとんど無い難病である。特に特発性肺線維症に対する有用な創薬標的分子はこれまでに見つかっていない。本研究の開発手法による予測で見出された特発性肺線維症に対する創薬標的候補分子は、実際にヒトの病態に近い SP-C ノックインマウスを用いた実験でその有効

性を示唆する結果が得られた。これは特発性肺線維症の創薬のみならず、繊維化の分子機序の解明にもつながる世界初の発見であり、非常に大きなインパクトがある。

<代表的な論文>

1. Iida, M., Iwata, M., and Yamanishi, Y., "Network-based characterization of disease-disease relationships in terms of drugs and therapeutic targets", *Bioinformatics*, 36, i516-i524, 2020.

概要: 疾患のオミックスデータや生体分子相互作用ネットワークの比較から、さまざまな病気に対して創薬標的分子を探索する機械学習手法を開発した。まず、2つの病気が分子間相互作用ネットワーク上でどれくらい似ているかを評価する方法を提案した。次に、病気間類似性と創薬標的分子の共通性の対応関係を学習し、創薬標的分子候補を探索する機械学習アルゴリズムをカーネル法の枠組みで開発した。実際に、代謝に関する分子間相互作用ネットワークを用いて、様々な病気の創薬標的分子の共通性を探索した。例えば、特発性肺線維症と2型糖尿病が創薬標的分子を共有する可能性が予測された。実際に、近年のマウスを用いた研究において、糖尿病の経口薬が特発性肺線維症の治療に効果があることが近年の文献で報告されており、この予測結果の妥当性を確認することができた。

2. Iwata, M., Yuan, L., Zhao, Q., Tabei, Y., Berenger, F., Sawada, R., Akiyoshi, S., Hamano, M., and Yamanishi, Y., "Predicting drug-induced transcriptome responses of a wide range of human cell lines by a novel tensor-train decomposition algorithm", *Bioinformatics*, 35, i191-i199, 2019.

概要: 最新のテンソル分解手法である tensor-train 分解アルゴリズムを用いた高階遺伝子発現データ解析手法を開発し、化合物応答遺伝子発現データの細胞ごとの未観測値や欠損値の補完に応用した。既存のテンソル分解手法よりも、データ補完性能において優れていることを確認した。特にテンソルの階数が高い化合物応答遺伝子発現データで、精度が高い(誤差が低い)ことを確認した。データ補完した化合物応答遺伝子データを用いて、様々な疾患の治療薬を予測する問題に応用した結果、データ補完しない場合に比べて精度が最大2倍向上することを確認した。

3. Tabei, Y., Kotera, M., Sawada, R., and Yamanishi, Y., "Network-based characterization of drug-protein interaction signatures with a space-efficient approach", *BMC Systems Biology*, 13(Suppl 2):39, 2019.

概要: 大規模な化合物のスクリーニングを可能にするため、機械学習の予測モデルを高速に学習するためのアルゴリズムの実装を行った。84,195,800次元の5,372,868タンパク質-化合物ペアのフィンガープリントのデータから化合物-タンパク質相互作用を予測するためのモデル構築を行った。データが次元と数の側面で巨大であるために、通常の機械学習アルゴリズムを適用すると、利用メモリーと学習時間の側面から適応困難である。そこで、我々は、化合物-タンパク質データを圧縮し、圧縮されたデータ上で機械学習モデルの学習アルゴリズムの開発を行った。トライ木によるデータ圧縮アルゴリズムを組み合わせ、予測モデルを学習することで、従来の手法では131GBのメモリーが必要なところを、我々の手法では24GBのメモリーで学習を行うことが可能となった。

§ 2 研究実施体制

(1) 研究チームの体制について

(1) 山西グループ

① 研究代表者: 山西 芳裕 (九州工業大学大学院情報工学研究院 教授)

② 研究項目:

・創薬標的検証用ツール化合物探索手法の開発と大規模予測

医薬ビッグデータに基づいて、対象疾患の創薬標的候補を予測する機械学習手法を開発する。さらに、その創薬標的の「確からしさ」の検証実験に利用可能なツール化合物を探索するインシリコ手法を開発する。様々ながん種および臓器線維症に関するマルチオミクスデータや分子ネットワークデータ、既承認薬、開発中止化合物、合成化合物、天然化合物など大規模な化合物の構造データや実験データを収集する。疾患データと化合物データの融合解析を行う統計手法や、多様なオミクス関連データを有効活用して生体分子や化合物を効率的にスクリーニングできる機械学習の手法を開発する。最終的に、特発性肺線維症および肺癌に対して見出された創薬標的分子候補を制御するツール化合物の候補をインシリコ予測する。

(2) 三森グループ

① 主たる共同研究者: 三森 功士 (九州大学大学病院 教授)

② 研究項目:

・がんに対するデータ解析と実験検証

肺癌に対するオミクスデータの解析や他のがん種のオミクスデータとの相関解析、ツールとなる可能性の高い化合物の選定を行う。細胞生存性、細胞毒性、アポトーシス誘導能など抗がん作用を *in vitro* で検証できる実験プラットフォームを構築し、予測したツール化合物の肺癌に対する効果を、実験的に検証する。創薬標的分子の候補をツール化合物で制御し、その抗腫瘍効果を検証する。

(3) 加藤グループ

① 主たる共同研究者: 加藤 明良 (大分大学医学部臨床薬理学講座 特任教授)

② 研究項目:

・線維症に対するデータ解析と実験検証

特発性肺線維症に対するオミクスデータの解析や他の関連疾患との相関解析、ツールとなる可能性の高い化合物を選定する。抗線維化作用を *in vitro* で検証できる実験プラットフォームを構築し、予測したツール化合物の特発性肺線維症に対する効果を、実験的に検証する。創薬標的分子の候補をツール化合物で制御し、その抗線維化作用を検証する。

(4) 田部井グループ

① 主たる共同研究者: 田部井 靖生 (理化学研究所革新知能統合研究センター ユニッティリーダー)

② 研究項目:

・アルゴリズムの開発と実装

膨大な数の化合物の構造データや実験データを効率的に処理するアルゴリズムを開発する。代表者が整備する化合物の大規模データを高速かつメモリ効率良く処理するため簡潔データ構造の技術の高度化を行う。また機械学習の手法とデータ圧縮技術を組み合わせて、大規模な化合物データから機械学習の予測モデルを高速に学習するためのアルゴリズムの実装を行う。特発性肺線維症や肺癌の創薬標的分子に対して、膨大な化合物のインシリコスクリーニングを行うための予測モデルの学習の際に、実装したアルゴリズムを用いる。

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

製薬会社など民間企業 5 社と共同研究契約を締結して、共同研究を実施した。
また、企業 1 社と共同研究を通じて、技術移転を行った。現在、別の民間企業 4 社から共同研究の依頼があり、共同研究契約締結に向けて、交渉中である。