

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 15 日現在

機関番号：17104

研究種目：基盤研究(B) (一般)

研究期間：2018～2020

課題番号：18H03334

研究課題名(和文) 医薬ビッグデータ解析による革新的AI創薬手法の開発と難治性疾患治療薬の発見

研究課題名(英文) AI-based drug discovery approach based on biomedical big data and its application to refractory diseases

研究代表者

山西 芳裕 (Yamanishi, Yoshihiro)

九州工業大学・大学院情報工学研究院・教授

研究者番号：60437267

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：本研究では、医薬ビッグデータと人工知能(AI)の基盤技術である機械学習を用いた創薬を実現するための情報基盤技術を構築する。医薬品、植物、食品成分などの化合物に関する情報、遺伝子、タンパク質、糖鎖など生体分子に関する情報、難治性疾患に関するオミックス情報やSNPなどのゲノム情報を基に、グラフ畳み込みニューラルネットワークや再帰的ニューラルネットワークの枠組みで、化合物の標的タンパク質プロファイルを予測するためのモデルを構築した。化合物の化学構造の適用範囲を考慮したアルゴリズムも構築した。最終的に、悪性リンパ腫と嚢胞性線維症に対して、医薬品候補化合物を網羅的に予測し、予測結果の一部を検証した。

研究成果の学術的意義や社会的意義

疾患治療に有用な化合物の同定は、人類の医療やヘルスケアにとって最重要課題である。現在でも有効な治療法が無い難治性疾患や希少疾患は多く、疾患に苦しむ患者に対する迅速な救済措置が必要である。しかしながら、最近の新薬開発は低迷しており、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いる。本研究では、医薬ビッグデータと人工知能(AI)の基盤技術である機械学習を用いた創薬「AI創薬」を提唱し、それを実現するための機械学習手法の研究開発を行なった。深層学習の予測モデルを構築し、医薬品候補の化合物を情報化学的にスクリーニングする技術基盤を構築することができた。

研究成果の概要(英文)：In this research, we build information infrastructure technologies to realize drug discovery using pharmaceutical big data and machine learning, a fundamental technology of artificial intelligence (AI). Based on information on compounds such as pharmaceuticals, plants, and food ingredients, information on biomolecules such as genes, proteins, and glycans, and information on intractable diseases such as omics data and genomic data including SNPs, we constructed models for predicting the target of compounds in the framework of graph convolutional neural networks and recursive neural networks. We also constructed algorithms that take into account the applicability domains of chemical structures of compounds. Finally, we comprehensively predicted drug candidate compounds for malignant lymphoma and cystic fibrosis, and validated some of the prediction results.

研究分野：バイオインフォマティクス

キーワード：機械学習 ビッグデータ 創薬 人工知能 難治性疾患

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

疾患治療に有用な化合物の同定は、人類の医療やヘルスケアにとって最重要課題である。現在でも有効な治療法が無い難治性疾患や希少疾患は多く、疾患に苦しむ患者に対する迅速な救済措置が必要である。しかしながら、最近の新薬開発は低迷しており、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いる。

その打開策として、近年の生命医学で大量に生み出されてきた様々な疾患に関するビッグデータを活用したデータ駆動型の創薬が期待されている。例えば、遺伝子やタンパク質に関するゲノム、トランスクリプトーム、プロテオームなどの多階層オミクスデータや、膨大な薬物に関する生理活性情報や薬物応答遺伝子発現情報が蓄積されつつある。しかしながら、そのような医薬ビッグデータを、創薬現場では十分に有効活用できていない。

近年の生命医学では、個々の疾患に関して網羅的オミクスデータが得られるようになり、個々の疾患の理解は進んでいるが、異なる疾患間の類似性を分子レベルで理解するには至っていない。問題として、個々の疾患のデータは公開されているものの、異なる疾患を横断解析するための情報基盤技術は整備されていないことがボトルネックになっている。

2. 研究の目的

本研究では、医薬ビッグデータと人工知能(AI)の基盤技術である機械学習を用いた創薬「AI創薬」を提唱し、それを実現するための情報基盤技術を構築する。多様な疾患に関する分子機序データ(遺伝子変異、発現異常遺伝子、異常パスウェイ、環境因子など)の横断解析から、疾患間の類似性を分子レベルで明らかにする。疾患類似性を基に、深層学習とマルチタスク学習を融合し、疾患に効能のある医薬品候補化合物を大規模に予測する機械学習アルゴリズムを開発する。数千万個の化合物をインシリコスクリーニングし、疾患に対して治療薬候補となる化合物を予測する。その疾患の中でも特に、難治性疾患である悪性リンパ腫と嚢胞性線維症に対する予測結果を実験で検証する。

3. 研究の方法

様々な疾患に関する病因遺伝子、配列変異、環境因子、診断マーカー、治療標的、患者の遺伝子発現情報、異常パスウェイなどの分子機序データを、OMIM, KEGG, GEOなどのデータベースや文献から整備する。これらの疾患の分子機序データをgenerative topographic map (GTM)や主成分分析で解析し、疾患間の類似性を表す疾患マップを構築する。疾患間の共通性を示す遺伝子、タンパク質、パスウェイを同定し、疾患の類似度を評価する。

化合物の標的タンパク質(オフターゲットを含む)を、化学構造情報、副作用情報、化合物応答遺伝子発現情報から予測する機械学習手法を開発する。化学構造からの予測においては、グラフ畳み込みニューラルネットワークや再帰的ニューラルネットワークを用いた非線形モデルやロジスティック回帰などの線形モデルを構築する。シングルタスク学習とマルチタスク学習の枠組みを検討し、クロスバリデーションなどの数値的シミュレーションでの性能評価を行う。さらに、化合物の化学構造の適用範囲を考慮したアルゴリズムを構築する。

疾患に対する化合物の効能の有無を予測する手法を実装する。疾患の分子機序の情報(遺伝子変異、発現異常遺伝子、疾患パスウェイ、環境因子、マーカーなど)で、疾患の類似性を評価する。疾患の類似性を考慮したモデルを開発する。提案手法を、合成可能かつ購入可能な化合物ライブラリーに適用し、解析対象の疾患に対する治療薬候補の化合物を予測する。

悪性リンパ腫・嚢胞性線維症に対して効能が新たに予測された化合物を購入して、実験的に検証する。悪性リンパ腫に対しては、候補化合物の抗がん作用を評価するため、細胞生存性、細胞毒性、アポトーシス誘導能を測定できる試薬などを用いて評価する。嚢胞性線維症の治療効果は、原因タンパク質CFTR Δ F508変異体の膜発現量および塩素イオンチャネル活性の回復で評価する。

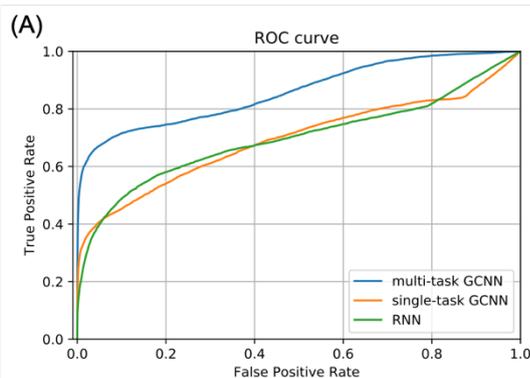
4. 研究成果

<研究の主な成果>

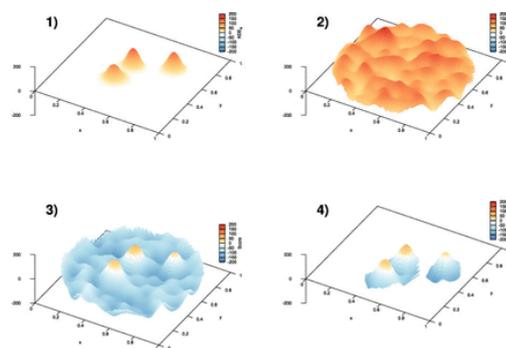
化合物、タンパク質、疾患に関する医薬ビッグデータを、インターネットや文献、データベースから適時収集して情報解析可能な形に整備した。承認された医薬品に関する情報を、KEGGなど様々なデータベースから得て、整備した。植物や食品成分の化合物に関する情報を、様々なデータベースから得て、整備した。また遺伝子、タンパク質、糖鎖に関する情報を、様々なデータベースから得て、整備した。難治性疾患に関するオミクス情報は、GEOなどのデータベースから得て、整備した。難治性疾患に関する病因遺伝子情報は、OMIMなどのデータベースから得て、整備した。難治性疾患患者のSNPなどのゲノム情報をGWASカタログなどの疾患特異的なデータベースから得て、整備した。

化合物の標的タンパク質プロファイルを構築した。グラフ畳み込みニューラルネットワークGCNNと再帰的ニューラルネットワークRNNの枠組みで、化合物の標的タンパク質プロファイル

を予測するためのモデルを構築した。シングルタスク学習とマルチタスク学習の枠組みを検討し、クロスバリデーションなどの数値的シミュレーションでの性能評価を行なった。予測精度は、ROC 曲線 (Receiver Operating Characteristic curve、受信者動作特性曲線) で評価した。ペプチド化合物の場合、マルチタスク学習での GCNN がシングルタスク学習での GCNN や RNN よりも高い精度を示した (右図参照)。ドッキングシミュレーションで、予測結果の検証を行った。例えば、ラクトリペプチドが ACE や DPP4 などのタンパク質のリガンド結合部に結合する様式を確認することができた。

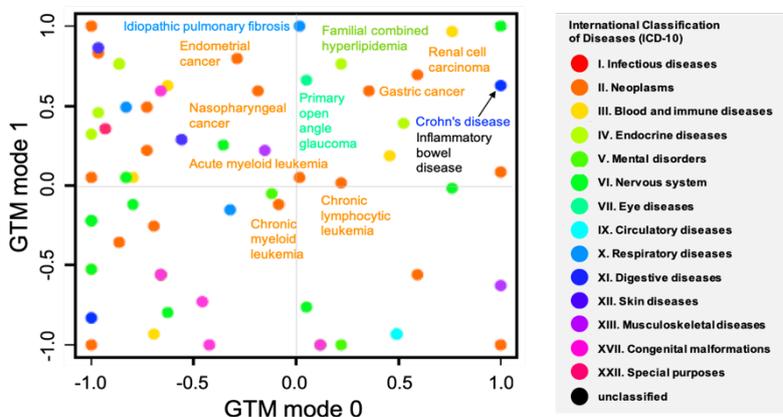


疾患の治療標的タンパク質に相互作用する医薬品候補化合物を予測するための機械学習手法を開発した。AI による化合物のヴァーチャルスクリーニングが期待されているが、学習データの中にある化合物の構造と似つかない化合物に関しては信頼性のある予測ができないという限界がある。そこで、Vanishing kernel というカーネル関数を用いて、予測モデルの適用範囲 (applicability domain) を定めて、化合物をインシリコスクリーニングする新しいアルゴリズムを開発した (右図参照)。予測精度は、深層学習のグラフ畳み込みニューラルネットワークの精度と大きな差はなかったが、計算時間は数百分の1と非常に計算効率が高いのが特徴である。ユーザーが指定するハイパーパラメータが1つだけなので、パラメータチューニングがやすく実用性が非常に高い。



1) KDE_a: KDE generated by actives. 2) KDE_i: KDE generated by inactives. 3) Score = KDE_a - KDE_i. 4) Score only inside the AD.

疾患のオミックスデータを用いて、疾患の分子機序の類似性を評価する手法を開発した。疾患患者の遺伝子発現プロファイルを疾患ごとに平均化し、疾患特異的遺伝子発現プロファイルを構築して、疾患間の相関を評価した。さらに、疾患間の共通性や疾患の特異性に関与する分子を抽出する方法を提案した。特に、糖鎖関連遺伝子に基づく疾患特異的遺伝子発現プロファイルを多様な疾患に対して構築し、generative topographic map (GTM) という方法を用いて、疾患間の関係性を視覚化した。副産物として、疾患を特徴付ける糖鎖構造を抽出することができた。



最後に、疾患に対して効能

を持つ化合物を予測する手法の開発を行った。疾患のオミックスデータや分子ネットワーク情報を用いて、疾患の分子機序の類似性を評価し、予測モデルに組み込んだ手法を開発した。疾患に対して効能を持つ化合物を予測する手法の開発を行った。対象疾患に対して、医薬品候補化合物を網羅的に予測した。抗がん作用を評価できる試薬や提案手法で予測された化合物を用意し、がんに対する考察や検証を行った。嚢胞性線維症の治療効果は、CFTR への作用などで評価するため、その評価系構築し、予測結果の一部を検証した。

<得られた成果の国内外における位置づけとインパクト>

疾患のオミックス情報を有効活用した医薬品候補化合物のスクリーニングは世界的にもあまり例がなく、非常に先駆的な研究成果を生み出すことができた。特に異なる疾患の横断解析を行い、疾患間の潜在的な相関に寄与する生体分子を抽出することは、疾患の病態メカニズムの理解にもつながることが期待でき、インパクトは大きい。

<今後の展望>

開発手法を様々な疾患に対して適用していく予定である。特に、新型コロナウイルス感染症に適用し、新しい治療薬候補の化合物の発見につなげていきたい。また単に医薬品候補化合物を予測するだけでなく、有用な創薬標的分子の発見をできるような手法の開発を行なっていきたい。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Fukunaga, I., Sawada, R., Shibata, T., Kaitoh, K., Sakai, Y., and Yamanishi, Y.,	4. 巻 39(1-2)
2. 論文標題 Prediction of the Health Effects of Food Peptides and Elucidation of the Mode-of-action Using Multi-task Graph Convolutional Neural Network	5. 発行年 2020年
3. 雑誌名 Molecular Informatics	6. 最初と最後の頁 e1900134
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/minf.201900134	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Akiyoshi, S., Iwata, M., Berenger, F., and Yamanishi, Y.	4. 巻 39(1-2)
2. 論文標題 Omics-based identification of glycan structures as biomarkers for a variety of diseases	5. 発行年 2020年
3. 雑誌名 Molecular Informatics	6. 最初と最後の頁 e1900112
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/minf.201900112	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Berenger, F., Zhang, K., and Yamanishi, Y.	4. 巻 11(1)
2. 論文標題 Cheminformatics and Structural Bioinformatics in OCaml	5. 発行年 2019年
3. 雑誌名 Journal of Cheminformatics	6. 最初と最後の頁 10
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s13321-019-0332-0	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Berenger, F. and Yamanishi, Y.	4. 巻 59(1)
2. 論文標題 A Distance-Based Boolean Applicability Domain for Classification of High Throughput Screening Data	5. 発行年 2019年
3. 雑誌名 Journal of Chemical Information and Modeling	6. 最初と最後の頁 463-476
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.8b00499	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計16件（うち招待講演 12件 / うち国際学会 3件）

1. 発表者名 Yamanishi, Y.
2. 発表標題 Data-driven drug discovery and medical treatment by machine learning
3. 学会等名 ACS Fall 2019 National Meeting & Exposition, Herman Skolnik Symposium (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yamanishi, Y.
2. 発表標題 Data-driven drug discovery and medical treatment by machine learning
3. 学会等名 The 6th Autumn School of Chemoinformatics in Nara 2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Fukunaga, I., Sawada, R., Shibata, T., Kaitoh, K., Sakai, Y., and Yamanishi, Y.
2. 発表標題 Prediction of Health Effects of Food Peptides and Elucidation of The Mode-of-action Using Multi-task Graph Convolutional Neural Networks
3. 学会等名 情報計算化学生物学会2019年大会
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：化合物の薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 第46回日本毒性学会学術年会 (招待講演)
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 第66回日本実験動物学会（招待講演）
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 第63回日本リウマチ学会総会・学術集会（招待講演）
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：化合物の薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 ファーマIT&デジタルエキスポ2019（招待講演）
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 機械学習によるデータ駆動型創薬とパスウェイ創薬
3. 学会等名 第92回日本薬理学会年会（招待講演）
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：化合物の薬効を予測するデータ駆動型アプローチ
3. 学会等名 第1回日本メディカルAI学会学術集会
4. 発表年 2019年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 第31回日本動物実験代替法学会（招待講演）
4. 発表年 2018年

1. 発表者名 山西芳裕
2. 発表標題 ディジーゾーム解析による疾患間の関連性理解と創薬応用
3. 学会等名 第12回メタボロームシンポジウム
4. 発表年 2018年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 第39回富山大学和漢医薬学総合研究所特別セミナー（招待講演）
4. 発表年 2018年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 新学術領域（研究領域提案型）「化学コミュニケーションのフロンティア」第1回若手シンポジウム（招待講演）
4. 発表年 2018年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 構造活性フォーラム2018（招待講演）
4. 発表年 2018年

1. 発表者名 山西芳裕
2. 発表標題 AI創薬：薬効や副作用を予測するデータ駆動型アプローチ
3. 学会等名 九州大学 - 理化学研究所 - 福岡市 三者連携シンポジウム
4. 発表年 2018年

1. 発表者名 Yamanishi, Y.
2. 発表標題 Data-driven drug discovery and repositioning by machine learning methods
3. 学会等名 ACS SkoInik Symposium（招待講演）（国際学会）
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	沖米田 司 (Okiyoneda Tsukasa) (90398248)	関西学院大学・理工学部・教授 (34504)	
研究分担者	谷 憲三郎 (Tani Kenzaburo) (00183864)	東京大学・定量生命科学研究所・特任教授 (12601)	
研究分担者	土方 康基 (Hijikata Yasuki) (80460856)	東京大学・医科学研究所・特任助教 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------