

令和 5 年 6 月 5 日現在

機関番号：14401

研究種目：若手研究

研究期間：2020～2022

課題番号：20K15422

研究課題名（和文）AIを活用したデータ駆動的研究による新規抗線維化薬の探索

研究課題名（英文）Identification of Novel Antifibrotic Drug by Data-Driven Research Using AI

研究代表者

野島 陽水（Nojima, Yosui）

大阪大学・数理・データ科学教育研究センター・特任講師（常勤）

研究者番号：30815717

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：初年度では化合物添加による細胞生存率を予測するAIの開発を目指した。2種類の交差検証法により評価を行った。次年度では前年度に構築したモデルにIPF患者のマルチオミクスデータおよび化合物情報を入力し、化合物ごとに細胞生存率を予測した。当初の計画ではIPF患者検体を購入しマルチオミクスデータを収集する予定であったが、新型コロナウイルス感染拡大の影響により検体入手が困難になったため、公共データからIPF患者肺のマルチオミクスデータを部分的に収集した。その結果、初年度で得られた精度を下回ったことから、遺伝子・タンパク質・代謝物等の複数のオミクスデータを統合することが精度向上に必須であることがわかった。

研究成果の学術的意義や社会的意義

IPFなどの難病はデータ量が少なく、IPF単独のデータでは人工知能による解析は不向きである。そこで本研究では、IPFと同じく細胞増殖によって発症し、情報が豊富にある癌のデータに着目した。単一の疾患では人工知能の入力データ量として不十分であっても、類似性のある疾患のデータを用いてカバーすることができれば、希少難病などの領域にも人工知能による解析アプローチが適用できると考えられる。

研究成果の概要（英文）：In the first year, I aimed to develop an AI for predicting cell survival rates through compound addition. The evaluation was conducted using two types of cross-validation methods. In the following year, I inputted multiomics data and compound information from IPF (Idiopathic Pulmonary Fibrosis) patients into the model constructed in the previous year to predict cell survival rates for each compound. Originally, our plan was to purchase IPF patient samples and collect multiomics data, but due to the impact of the COVID-19 pandemic, it became difficult to obtain the samples. Therefore, I partially collected multiomics data from IPF patient lungs through public databases. As a result, it was found that integrating multiple omics data, such as genes, proteins, and metabolites, is essential for improving accuracy, as it fell below the accuracy obtained in the first year.

研究分野：生命科学、バイオインフォマティクス、計算生物学

キーワード：人工知能 AI 機械学習 IPF ドラッグリポジショニング 計算生物学 創薬探索 マルチオミクス

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

特発性肺線維症 (Idiopathic pulmonary fibrosis; IPF) は、遺伝的要因および環境的要因によって肺に傷ができ、その修復過程では肺線維芽細胞から過剰にコラーゲンが産生され、結果として間質が厚くなり呼吸困難となる難病である。5 年生存率 30% と予後不良な疾患であり、2 剤の抗線維化薬が承認されているものの、いずれも進行速度を遅くする効果しかないため、既存薬から発展させた治療薬が望まれている。

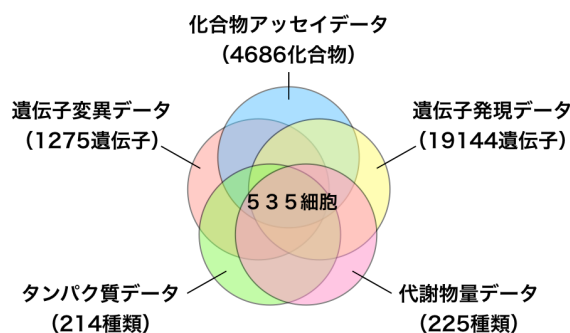
IPF 発症には肺線維芽細胞のアポトーシス回避および増殖が関与し、結果として肺線維芽細胞におけるコラーゲン合成が継続的に行われる (D Alvarez et al., 2017 Am J Physiol Lung Cell Mol Physiol, C Huang et al., 2017 J Biol Chem)。したがって、根本的な治療には肺線維芽細胞のアポトーシス回避および増殖を阻害することが有効であると考えられる。また、線維化は癌化と密接に関連しており、IPF と肺癌のリスクファクターは共通していることが知られている (B Ballester et al., 2019 Int J Mol Sci)。研究代表者による先の検討で、肺癌と IPF を含めた 3 つの呼吸器疾患の間で遺伝子発現パターンを比較したところ、肺癌は他の呼吸器疾患よりも IPF の遺伝子発現パターンに近いことが明らかとなり、癌化と線維化は類似の発症機序を有することが示唆された。これらのことから、癌のデータから新たな線維化治療薬を提案できる可能性は高い。

2. 研究の目的

本研究では、ヒト癌細胞株に関するビッグデータを人工知能により解析し、肺線維芽細胞の増殖を抑制する化合物の提案を目的とする。

3. 研究の方法

DepMap には、535 細胞株に対して 4686 化合物を添加し細胞生存率を算出したアッセイデータが格納されており、この全ての細胞株に遺伝子発現、遺伝子変異、タンパク質発現、代謝物量といった複数のオミクスデータが紐付いている (下図)。



DepMap データの概要

つまり、250万通りのアッセイデータとそれに紐づくマルチオミクスデータが格納されており、そのうち肺由来細胞株の格納数は最も多い。DepMap に格納されているマルチオミクスデータと化合物の記述子を連結させ、訓練データとする。マルチオミクスデータのうち、発現データは数値として、遺伝子変異データは2進コード化して使用する。記述子は、化合物の構造から記述子計算ソフトを用いて算出する。教師データには化合物を細胞に添加した際の細胞生存率を用いて、化合物添加による細胞の生存率を予測するAIの構築を目指す。学習アルゴリズムには主として、Deep Neural Network (DNN)を用いる。その他に、Random forest や Support vector machine などの機械学習アルゴリズムも比較検討する。モデルの精度評価は、実測細胞生存率と予測細胞生存率を用いて、Root Mean Squared Error 等を算出して行う。開発したAIにIPF患者肺から樹立した初代培養細胞のマルチオミクスデータと化合物情報を入力し、IPF治療に有効な候補化合物を提案する。最終的には、実験的検証により化合物の薬理効果を確認する。

4. 研究成果

初年度では、化合物添加による細胞生存率を予測するAIの開発を目指した。具体的には、公共がん細胞株データベースのマルチオミクスデータと化合物情報を説明変数、化合物添加による細胞生存率を目的変数として、複数の学習アルゴリズムを用いてモデルを構築した。また、k-分割交差検証法およびleave-one-out交差検証法により評価を行った。

次年度では、前年度に構築したモデルにIPF患者のマルチオミクスデータおよびChEMBLの化合物情報を入力し、化合物ごとの細胞生存率の予測を試みた。当初の計画ではIPF患者の肺線維芽細胞の初代培養細胞を購入し本研究においてマルチオミクスデータを収集する予定であったが、新型コロナウイルス感染拡大の影響により検体入手が困難になったことから、公共データか

ら IPF 患者肺のマルチオミクスデータを探索・収集し、これを用いることにした。

最終年度の当初の計画では、市販の IPF 患者の肺線維芽細胞の初代培養細胞を購入し、薬理効果の検証を行う予定であったが、新型コロナウイルス感染拡大の影響により検体入手が困難な状況となった。また、令和3年度の当初の計画においても、IPF 患者の肺線維芽細胞の初代培養細胞を購入し、マルチオミクスデータを収集する予定であったが、上記と同様の理由により検体入手が困難になったことから、公共データから IPF 患者肺のマルチオミクスデータを探索・収集し、これを用いることにした。しかし、公共オミクスデータから得られる説明変数の数（遺伝子やタンパク質数など）は、初年度のモデル構築時の説明変数の数を十分にカバーしていなかったため、公共オミクスデータに合わせたモデル構築を再度実施した。その結果、初年度に交差検証で得られた精度を下回ったことから、遺伝子・タンパク質・代謝物等の複数のオミクスデータを統合することが精度向上に必須であることがわかった。

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 9件 / うち国際共著 0件 / うちオープンアクセス 9件）

1. 著者名 Nojima Yosui, Takeda Yoshito	4. 巻 volume 370
2. 論文標題 Omics Data Analysis Tools for Biomarker Discovery and the Tutorial	5. 発行年 2021年
3. 雑誌名 Springer Proceedings in Mathematics & Statistics	6. 最初と最後の頁 266 ~ 273
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-981-16-4866-3_19	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 Asakura Nobuhiko, Nakamura Naotoshi, Muroi Atsushi, Nojima Yosui, Yamashita Taro, Kaneko Shuichi, Ikeda Kazuki, Koshikawa Naohiko, Suzuki Takashi	4. 巻 22
2. 論文標題 Expression of Cancer Stem Cell Markers EpCAM and CD90 Is Correlated with Anti- and Pro-Oncogenic EphA2 Signaling in Hepatocellular Carcinoma	5. 発行年 2021年
3. 雑誌名 International Journal of Molecular Sciences	6. 最初と最後の頁 8652 ~ 8652
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/ijms22168652	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 Futami Yu, Takeda Yoshito, Koba Taro, Narumi Ryohei, Nojima Yosui., et al	4. 巻 in press
2. 論文標題 CD14 and lipopolysaccharide-binding protein as novel biomarkers for sarcoidosis by proteomics of serum extracellular vesicles	5. 発行年 2022年
3. 雑誌名 International Immunology	6. 最初と最後の頁 in press
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/intimm/dxac009	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 Hirano Hidekazu, Abe Yuichi, Nojima Yosui, Aoki Masahiko, Shoji Hirokazu, Isoyama Junko, Honda Kazufumi, Boku Narikazu, Mizuguchi Kenji, Tomonaga Takeshi, Adachi Jun	4. 巻 12
2. 論文標題 Temporal dynamics from phosphoproteomics using endoscopic biopsy specimens provides new therapeutic targets in stage IV gastric cancer	5. 発行年 2022年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 open access
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-022-08430-7	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 Natsume-Kitatani Yayoi., et al.	4. 巻 -
2. 論文標題 Data-driven patient stratification and drug target discovery by using medical information and serum proteome data of idiopathic pulmonary fibrosis patients	5. 発行年 2022年
3. 雑誌名 Research Square	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.21203/rs.3.rs-405195/v2	査読の有無 無
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 Nojima Yosui, Takeda Yoshito, Maeda Yohei, Bamba Takeshi, Fukusaki Eiichiro, Itoh Mari N., Mizuguchi Kenji, Kumanogoh Atsushi	4. 巻 10
2. 論文標題 Metabolomic analysis of fibrotic mice combined with public RNA Seq human lung data reveal potential diagnostic biomarker candidates for lung fibrosis	5. 発行年 2020年
3. 雑誌名 FEBS Open Bio	6. 最初と最後の頁 2427 ~ 2436
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/2211-5463.12982	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 Allendes Osorio Rodolfo S., Johan Nystrom-Persson, Nojima Yosui, Kosugi Yuji, Mizuguchi Kenji, Natsume-Kitatani Yayoi	4. 巻 6
2. 論文標題 Panomicon: A web-based environment for interactive, visual analysis of multi-omics data	5. 発行年 2020年
3. 雑誌名 Heliyon	6. 最初と最後の頁 e04618 ~ e04618
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.heliyon.2020.e04618	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 Esaki Tsuyoshi, Horinouchi Takaaki, Natsume-Kitatani Yayoi, Nojima Yosui, Sakane Iwao, Matsui Hidetoshi	4. 巻 20
2. 論文標題 Estimation of relationships between chemical substructures and antibiotic resistance-related gene expression in bacteria: Adapting a canonical correlation analysis for small sample data of gathered features using consensus clustering	5. 発行年 2020年
3. 雑誌名 Chem-Bio Informatics Journal	6. 最初と最後の頁 58 ~ 61
掲載論文のDOI (デジタルオブジェクト識別子) 10.1273/cbij.20.58	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 Koba Taro, Takeda Yoshito, Narumi Ryohei, Shiromizu Takashi, Nojima Yosui., et al.	4. 巻 7
2. 論文標題 Proteomics of serum extracellular vesicles identifies a novel COPD biomarker, fibulin-3 from elastic fibres	5. 発行年 2021年
3. 雑誌名 ERJ Open Research	6. 最初と最後の頁 00658 ~ 2020
掲載論文のDOI (デジタルオブジェクト識別子) 10.1183/23120541.00658-2020	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 Nojima Yosui, Aoki Masahiko, Re Suyong, Hirano Hidekazu, Abe Yuichi, Narumi Ryohei, Muraoka Satoshi, Shoji Hirokazu, Honda Kazufumi, Tomonaga Takeshi, Mizuguchi Kenji, Boku Narikazu, Adachi Jun	4. 巻 21
2. 論文標題 Integration of pharmacoproteomic and computational approaches reveals the cellular signal transduction pathways affected by apatinib in gastric cancer cell lines	5. 発行年 2023年
3. 雑誌名 Computational and Structural Biotechnology Journal	6. 最初と最後の頁 2172 ~ 2187
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csbj.2023.03.006	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 -

〔学会発表〕 計2件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 今井 綾, 野島 陽水, 八尾 良司, 鈴木 貴
2. 発表標題 シングルセルRNA-Seq解析による家族性大腸腺腫症悪性化メカニズムの解明
3. 学会等名 応用数理学会 第18回研究部会連合発表会
4. 発表年 2022年

1. 発表者名 野島 陽水
2. 発表標題 オミクスデータを活用した肺線維症バイオマーカー探索に関する研究
3. 学会等名 応用数理学会 第17回研究部会連合発表会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

大阪大学 研究者総覧
<https://rd.iai.osaka-u.ac.jp/ja/738018fe46601a49.html>

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
--	---------------------------	-----------------------	----

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
---------	---------