

令和 4 年 6 月 2 日現在

機関番号：12601

研究種目：基盤研究(A) (一般)

研究期間：2018～2021

課題番号：18H04162

研究課題名(和文) がん分子標的薬の合理的投与設計手法の構築

研究課題名(英文) Optimization for usage of molecular targeted drugs on cancer therapy

研究代表者

鈴木 洋史 (Suzuki, Hiroshi)

東京大学・医学部附属病院・教授

研究者番号：80206523

交付決定額(研究期間全体)：(直接経費) 34,200,000円

研究成果の概要(和文)：本研究において、既存の大規模データベースを用い、細胞の分子情報を用いることで薬物感受性を濃度依存性も含め一定程度予測できるモデルを機械学習にて確立した。モデルの精度検証過程において、環境によって細胞の薬物感受性が異なることが実験的に確認され、環境差を考慮した改変・実装が今後の課題として見いだされた。ただし現行モデルの内部においても、定性的に既存の生物学的知見と一致する要素が学習されていることを確認しており、感受性規定因子の定性的な因子探索に活用できる。これらの結果から、がん細胞の分子基盤情報に基づいて合理的な抗がん剤を選択するためのモデルとして活用可能と考えられる。

研究成果の学術的意義や社会的意義

抗がん剤治療の最適化に関して様々な取り組みが行われているものの、十分な治療効果が得られる割合は必ずしも高くなく、種々の副作用出現により治療を断念せざるを得ない場合も少なくない。本研究では細胞の薬物感受性を予測するモデル構築により、適切な薬物選択を可能とする基盤構築を目指した。モデルの定量的信頼度向上の取り組みは今後の課題であるが、薬剤選択に関連する定性的因子を抽出可能なモデルの構築に成功した。がん細胞および正常細胞への感受性予測を併せることで、薬理効果および一部の副作用発現リスクも考慮した合理的な薬剤選択に寄与する研究成果である。

研究成果の概要(英文)：In this study, we constructed a machine learning model in which cell survival curves under anti-cancer drugs exposures can be calculated from molecular information in cells by utilizing large-scale public database. The quantitative accuracy of this model is, however, limited since we experimentally observed that the cellular sensitivity was varied by experimental environments. Therefore, a future challenge for this study is to construct a model in which environmental difference can be calibrated. Although there remains such a challenge, we confirmed that our current model could provide qualitative factors related to cellular sensitivity. Thus, our constructed model in this study can be used as a base model to reasonably select anti-cancer drugs by using molecular information about cancer cells.

研究分野：システム薬理学

キーワード：細胞毒性 機械学習 システム薬理学 抗がん剤

1. 研究開始当初の背景

がんは日本人の死因の3割以上を占めており、増加傾向が続いているが、近年種々の分子標的薬が開発された結果、薬物療法も大きく変化してきた。中でもキナーゼ阻害薬は、選択肢が多く治療効果も強いため、分子標的薬の中でも中心的な役割を果たしている。一方、キナーゼ阻害薬の臨床使用拡大に伴って副作用発現が大きな問題となってきた。キナーゼ阻害薬を中心とする分子標的治療薬の副作用は、その発症スペクトルが従来の化学療法剤と大きく異なり、経験的な対処法が必ずしも有効でない。その結果、投与中断を余儀なくされる例が多く、結果として治療効果が得られなくなるため、臨床上大きな課題となっている。キナーゼ阻害薬による副作用は、開発時に想定されたキナーゼ以外のいわゆるオフターゲット阻害に起因することが示唆されており、実際にキナーゼ阻害薬とキナーゼの網羅的親和性測定に関する報告により、ほぼ全てのキナーゼ阻害薬にオフターゲット阻害が起こりうる事が明らかとなっている(Karaman MW *et al.*, *Nat Biotech* 2009)。これらの分子基盤情報に基づく副作用発現や治療効果の定量的予測は、前述の臨床課題を克服し治療効果の最大化をもたらすと考えられるが、研究開始時点においては、確立された方法論は存在していなかった。

2. 研究の目的

上述背景のもと、副作用・治療効果発現の分子機構解析を進め、副作用を軽減しつつ治療効果を維持するための、合理的な薬物治療法の提唱を目的として、本研究が計画されている。研究代表者らは、研究開始時点において、一部のキナーゼ阻害と副作用の関連に関して、毒性標的分子の同定と副作用軽減手法の提唱を進めてきており、本研究では、これらの研究の発展的展開として、臨床使用されている一連のキナーゼ阻害薬を含む抗がん剤による薬理効果および副作用発現を分子基盤に基づいた予測が可能となるような基盤システムの構築を目指すこととした。

3. 研究の方法

分子基盤と種々の抗がん剤の薬理効果の連関をモデル化する場合、薬剤が各種細胞に与える細胞障害性を予測するモデルが基盤として利用できると期待される。そこで、本研究では、各種抗がん剤が細胞に与える毒性を定量的に記述可能なモデルの構築を目指すこととした。このモデル構築にあたり、既存の個別の細胞内分子挙動に関する多種多様な知見の背後に数式を実装する知識ベースのモデル化アプローチと、細胞内分子挙動に関する個別知見は考慮せずに、細胞内の分子状態と薬剤暴露による影響に関する実測データから、データ駆動的にモデル化するアプローチが考えられる。いずれの手法であっても、分子状態と細胞障害性を連関できるため、本研究での目的と合致しているが、前者の知識ベースモデル化に関しては、背後の数式実装の過程を精密に行うために必要となる情報であっても、実験的に測定不可能なものも多く存在することが予見されたため、本研究においては、データ駆動型のモデル構築を機械学習にて行うこととした。具体的には、以下の通りである。

(1) 機械学習を用いた抗がん剤曝露下における細胞生存予測モデルの構築

A. キナーゼ阻害薬に関する細胞障害性予測モデルの構築

未知の感受性規定要因の探索に将来的に応用することを念頭に、既知の感受性規定要因に限定せず、がん細胞の分子基盤に関する情報として網羅的な遺伝子発現/変異情報を用い、薬物情報については一般的な分子量や構造情報も用いることで、薬物曝露下のがん細胞の感受性を予測可能か検証することとした。従来よりキナーゼ阻害薬に着目した研究を進めてきた背景もあり、まずはキナーゼ阻害薬に関する感受性予測に焦点を絞った。キナーゼ阻害薬については、各キナーゼと阻害薬のアフィニティが薬理効果に大きな影響を与えることが想定されたため、モデル構築の際には、薬物と各種キナーゼに対する網羅的な親和性プロファイルも薬物情報として用いた。これらの細胞・薬物の情報に対応する細胞感受性データとして、Genomics of Drug Sensitivity in Cancer (GDSC) プロジェクトのデータを用いた。このプロジェクトでは、がん細胞の遺伝子発現量・変異情報、および、がん細胞の *in vitro* 薬物感受性を大規模に評価しており、大規模な感受性情報が入手可能である。GDSC1 として公開されているデータとして、約 1000 種類の細胞に対して、約 70 種類のキナーゼ阻害薬が与える細胞障害性データを、薬物濃度依存的な細胞生存率として *in vitro* にて測定した結果が入手可能であった (1000×70 の全ての組み

合わせが実測されている訳ではない)。そこで、入手可能なデータを用い、細胞感受性の代表値である EC50 値を予測できるよう教師あり機械学習の手法にてモデル構築した。

B. キナーゼ阻害薬以外の薬剤を含む抗がん剤全般の細胞障害性予測モデルの構築

続いて、上記(A)でのモデルをより定量的に拡張すること、かつ、対象となる薬物を拡大することとした。GDSC1 として登録されている薬物は、キナーゼ阻害薬に限定すると 70 種類程度であるが、その他の抗がん剤を含めると、薬 300 種類の薬剤に関する感受性実測データが登録されている。また、in vitro 細胞障害性評価試験における細胞生存率評価を行った生データが入手可能であり、感受性代表値としてしばしば用いられる EC50 値のみならず、濃度依存的な細胞生存率プロファイルそのものが入手可能である。そこで、細胞情報や薬物情報としては、上記(A)と同様のデータを用い、そこに薬物濃度の情報を与えた際に、細胞生存率を出力(予測)するモデルの構築を行った。

(2) 構築したモデルの汎化性能の検証

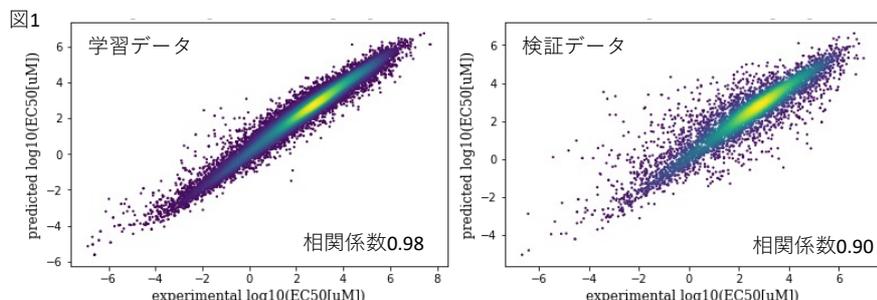
学習精度に影響を与える要因として、全学習データのうち各細胞あるいは各薬物の情報量が占める割合、感受性の大小、実験環境による感受性差が想定される。そこで、細胞あるいは薬物毎に、情報量の多寡および報告値の平均 EC50 値の大小に従って薬物および細胞をカテゴリ分類した上で、各カテゴリから代表的な細胞および薬物を選択し、実際に感受性を測定し、データベース報告あるいはモデル予測値との比較することで、各種バイアスを低減させた状況での性能評価を試みた。なお、薬物の選択は、細胞の遺伝子発現プロファイルによる応答性の違いが鋭敏に反映されることを期待し、キナーゼ阻害薬から選択することとし、キナーゼ阻害剤 10 種類について、11 種類のがん細胞株に対して、細胞生存への影響が観察できる濃度範囲において、濃度依存的な細胞生存率の評価を行った。

4. 研究成果

(1) 機械学習を用いた抗がん剤曝露下における細胞生存予測モデルの構築

A. キナーゼ阻害薬に関する細胞障害性予測モデルの構築

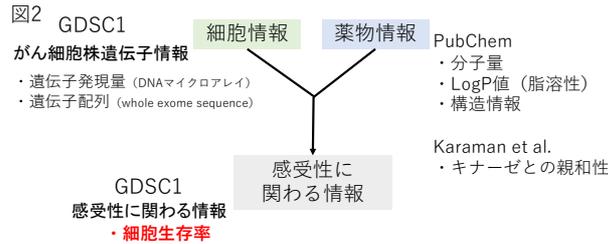
モデル骨格として、細胞情報および薬物情報を入力すると、感受性を出力する形式を設定した。細胞情報として、19,834 遺伝子の発現量、17,003 遺伝子についてその遺伝子内の変異数を用い、また、薬物情報としての分子量、オクタノール/水-分配係数、2,048 bit に変換した Morgan フィンガープリント、及び 442 種のキナーゼとの結合定数を、それぞれ入力情報として用いた。出力情報としては、GDSC1 にて報告されている EC50 実測推定値を用いた。なお、キナーゼとの結合定数は、過去に報告されている文献での報告値(Karaman MW *et al.*, *Nat Biotech* 2009 など)を参照したが、GDSC1 にて感受性が評価されているキナーゼ阻害薬約 70 種類のうち、36 種類のみデータが入手可能であった。そのため、約 900 種類の細胞と 36 種のキナーゼ阻害剤をモデル構築の対象とした。感受性が報告されている薬物-細胞の組み合わせのうち、8 割を学習データ、2 割を検証データに割り付け教師あり学習を行った。モデルの構築は、Random Forest Regression, Support Vector Regression, Multi-layer Perceptron の 3 つの手法にて行った。また、比較対象のベースラインモデルとしては、細胞あるいは薬物を index 化させたデータに対して線形回帰させたモデルを構築した。その結果、Random forest regression モデル、Support vector regression モデルでは、学習データにおける予測 EC50 値対実測 EC50 値の相関係数は 0.85 付近であり、ベースラインモデルでの相関係数(0.80)とよりやや良好な程度であった。一方で、Multi-layer perceptron モデルでは、学習データにおける予測対実測の相関係数が 0.98、



検証データでの相関係数は 0.90 (ベースラインモデルでの相関係数は 0.78) と極めて良好な結果となった(図1)。以上より、Multi-layer perceptron モデルによって、細胞の薬物感受性に関する情報良好に抽出可能であると考えられた。

B. キナーゼ阻害薬以外の薬剤を含む抗がん剤全般の細胞障害性予測モデルの構築

上記の結果を受けて、細胞感受性情報より定量的に拡張すること、かつ、対象となる薬物を拡大する試みをすすめた。対象薬物は、GDSC1 に登録されている約 300 種類全てを用いた。



これら薬物の大部分は、上記(A)で入力情報として用いたキナーゼとの結合定数が入手不能であったため、入手不能な部分の結合定数は0として学習を行った。

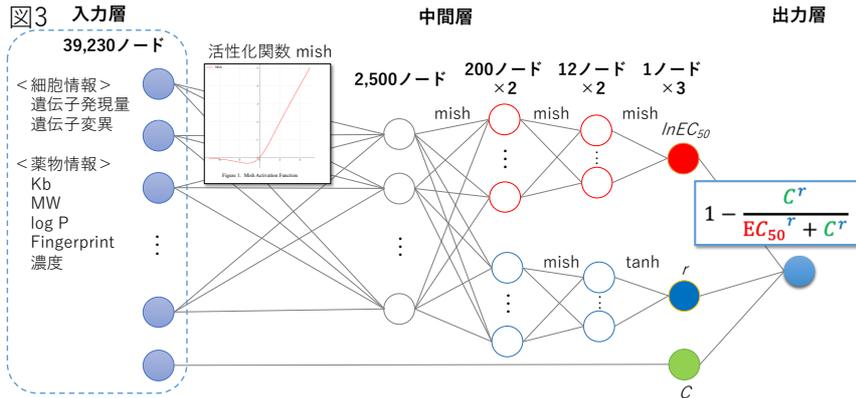
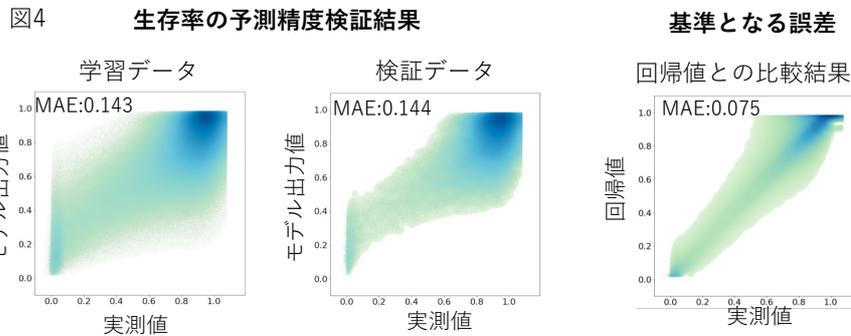


図2に示すように、細胞生存率を出力とする様々な抗がん剤に利用可能なモデルの構築を目指すこととし、Multi-Layer Perceptron のモデル形式を基盤として、かつ、細胞生存率を Hill-EC50 形式に基づき算出する図3のようなモデル骨格に対し、生存率を学習させた。薬物-細胞の組み合わせにおける細胞生存率データを個別に Hill-EC50 形式に回帰した場合の回帰生存率と実測生存率の誤差を比較対象とすると、学習データおよび検証データは、何れも実測値の回帰にて得られた誤差の2倍程度であり、十分な精度を有しているモデルが構築出来たことが確認できた。

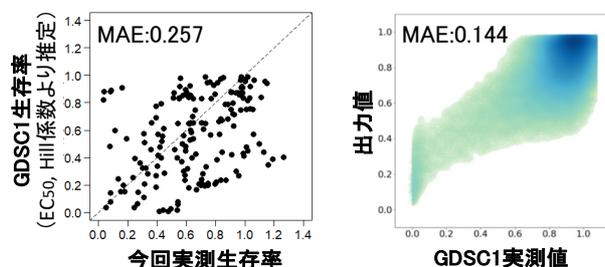


(2) 構築したモデルの汎化性能の検証

環境差、薬物感受性差、情報量差の検証を評価することを目的に、当研究室においても、実際に一定数の細胞-薬物の組み合わせについて細胞生存率評価を行い、モデル出力と比較することとした。この際、様々な細胞に対する感受性が高い、すなわち、報告データにおける EC50 値の対数平均値が低い薬物から、感受性が低い薬物までを偏り無く、10 種類 (Selumetinib、Midostaurin、BI-2536、Quizartinib、Axitinib、Erlotinib、Ruxolitinib、Doramapimod、Imatinib、Dasatinib) 選定した。また、GDSC1 データベースにて報告されている情報量 (すなわち、各薬剤について測定が実施されている細胞数) が少ない薬物から多い薬物まで、偏りの無い選別を行った。また、細胞についても、感受性の高低や情報量の多寡に応じて、偏りが生じないように 11 種類 (KYAE-1、HSC-3、ECC12、A2058、KP-4、VMRC-LCD、KYSE70、HLE、MKN7、SCC-3、SKM-1) を選定した。その結果、GDSC1 にて報告がある感受性評価されている組み合わせに関して、今回の実測生存率と報告生存率を対比すると、上記(1B)で構築したモデルの予測誤差を上回る差異 (環境差) が認められた (図5)。また、GDSC1 での実測報告がない薬物-細胞の組み合わせについて、モデル出力の生存率と比較すると、環境差と同程度の誤差が生じていることが確認された。この

ことから、少なくとも環境差を上回る予測誤差は発生していないことが確認できた。今後の課題としては、環境差に対する補正項を導入するなどといったモデルの改善が必要と考えられた。環境差補正に関する更なる検討は必要ではあるが、一定の環境内では一定の精度を有するモデルが構築できていることから、機械学習によってモデルが「学習した内容」を解析することで、定性的ではあるが感受性を規定する因子を抽出できるモデルとなっていることが期待された。そこで、モデルの各入力因子に対する出力の影響を評価する解析を進めたところ、一例としてEGFR阻害剤とEGFR遺伝子の発現量の多寡がモデル出力すなわち生存率に関連していることが見いだされるなど、構築したモデルにおいて、生物学的知見が適切に包含されており、分子基盤にリンクしたものとなっていることが確認された。

図5



(3) まとめ

本研究において、細胞内分子情報を用いることで、一定の環境下における細胞の薬物感受性を濃度依存性も含め、一定程度予測できるモデルが確立された。一方で、環境によって細胞の薬物感受性が異なることが実験的に確認され、モデル改良のさらなる課題として環境差を考慮したモデル改変・実装が見いだされた。ただし、モデル内部で学習された内容の評価すると、定性的に既存の生物学的知見と一致する内容が確認はされており、毒性発現因子の定性的あるいは半定量的な因子探索に活用できるモデルが構築できていると考えられた。なお、今回のモデル構築では、多岐にわたるがん細胞の薬物感受性を教師有り機械学習の学習データとして用いており、様々な臓器由来の細胞を包含しているデータであるため、各種抗がん剤の正常細胞への影響評価にも適応できることが期待される。抗がん剤の薬理効果のみならず、正常細胞への影響といった副作用予測にも応用可能と考えられることから、本研究での成果は、がん細胞の分子情報から合理的な薬物選択を行う過程に寄与するものだと考えている。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1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 Shinichi Uchida, Yoshiyuki Asai, Yoshiaki Kariya, Kunichika Tsumoto, Hiroshi Hibino, Masashi Honma, Takeshi Abe, Fumiaki Nin, Yasutaka Kurata, Kazuharu Furutani, Hiroshi Suzuki, Hiroaki Kitano, Ryuji Inoue, Yoshihisa Kurachi	4. 巻 69
2. 論文標題 Integrative and theoretical research on the architecture of a biological system and its disorder	5. 発行年 2019年
3. 雑誌名 The Journal of Physiological Sciences	6. 最初と最後の頁 443-451
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s12576-019-00667-8	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1件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 Hiroshi Suzuki
2. 発表標題 Basic and clinical studies toward the precision medicine
3. 学会等名 日本薬理学会年会サテライトシンポジウム（招待講演）（国際学会）
4. 発表年 2019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	本間 雅 (Honma Masashi) (60401072)	東京大学・医学部附属病院・講師 (12601)	
研究分担者	苅谷 嘉頭 (Kariya Yoshiaki) (20633168)	東京大学・医学部附属病院・助教 (12601)	

6. 研究組織（つづき）

	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	池淵 祐樹 (Ikebuchi Yuki) (20645725)	東京大学・医学部附属病院・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