



分野：生命科学・医学系

キーワード：シミュレーション、ソフトウェア、患者固有モデル、創薬

イン・シリコ患者固有モデルでがんの予後と薬剤応答を予測

—細胞シミュレーションによる疾患分類法の開発—

【研究成果のポイント】

- ◆ 生化学反応過程を説明する文章を数式・数理モデルに自動変換する新モデル構築手法「Text2Model」を開発
- ◆ 数学の知識がなくとも生物学の知識だけで簡単な数理モデルの構築を実現
- ◆ イン・シリコの分子活性動態からがんの予後と応答薬剤を予測
- ◆ 個人の遺伝子特性を反映した患者固有モデル「Pasmopy」ソフトウェアの構築により創薬や個別化医療を加速

❖ 概要

大阪大学蛋白質研究所 細胞システム研究室 岡田眞里子教授らの研究グループは、同大学大学院理学研究科の大学院生 井元宏明さん(博士後期課程)と山城紗和さん(博士前期課程)を中心に、イン・シリコ^{※1}の患者固有モデル(patient-specific model)^{※2}の構築手法を開発・公開し、臨床の遺伝子発現データからトリプルネガティブ乳がん^{※3}における個々の患者の予後と応答薬剤の予測に成功しました。

これまでがん治療においては、がんの大きさ、浸潤の程度、ある特定の疾患遺伝子マーカーの有無などの指標に基づき治療法が選択されてきました。しかし、これまでの分類法では患者の予後に大きな違いが生じることから、個々の患者の多様な遺伝子情報に基づいた新たな分類法の開発が急務となっています。そこで、本研究では、がんの層別化(たくさんのがんのデータを遺伝子の特徴や薬剤応答によってグループ分けすること)や個別化医療を目標とした患者固有モデルの構築を目指しました。

そのために、臨床の公共データベース^{※4}から取得したがん細胞株およびがん患者のRNAシーケンス(網羅的な遺伝子発現)データと細胞シミュレーション^{※5}を組み合わせ、計算により得られたイン・シリコの分子活性の動態から各患者の予後と治療標的を予測する計算手法を開発しました。この方法を用いると、少数の細胞株の実験データを数理モデルに学習させるだけで、多数のがん患者検体由来のRNAシーケンスデータのみを入力することで、各患者の予後と、患者ごとに適した薬剤の探索が患者固有シミュレーションのみから予測できるようになるため、実験研究を減らして創薬探索研究のデジタル化を手助けすることができます。

さらに、本研究では、細胞シミュレーションの利便性を高めるため、多くのがんで重要な働きを担う膜受容体シグナル伝達系のネットワークを例として、分子間の結合解離・酵素反応・分子局在・分解などの生化学反応の文献情報(テキスト)を連立常微分方程式モデルへと変換する新たな計算手法「Text2Model」を構築しました。また、本解析では、感度解析^{※6}という手法により、予後の悪いトリプルネガティブの患者群において、EGFR阻害剤^{※7}への感受性が低いことを予測し、それを細胞実験データにより検証しました。

このような数理モデルを用いた細胞シミュレーション技術は、疾患メカニズムの同定や定量的解析において優れており、ビッグデータに基づく分類に優れた人工知能(AI)と相補的に創薬研究に用いられることが期待されています。本研究における患者固有モデリング基盤 Pasmopy (Patient-Specific Modeling in Python)は、モデルの構築と個別化に必要な機能を集積し、オープンソースのソフトウェアとして公開されています(<https://github.com/pasmopy/pasmopy>) (図1)。本研究成果は、国際科学誌『Science』に、3月11日(金)(日本時間)に公開されます。

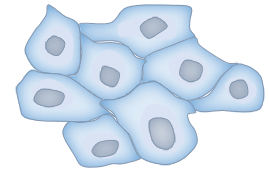
生化学反応の説明文

```
L binds R <--> LR
LR phosphorylates K --> pK
LR is degraded
pK phosphorylates TF --> pTF
pTF transcribes mRNA
mRNA is translated into P
P dephosphorylates pK --> K
```

1. テキスト から 数式への自動変換

2. 細胞株データによるパラメータ推定

細胞株データ

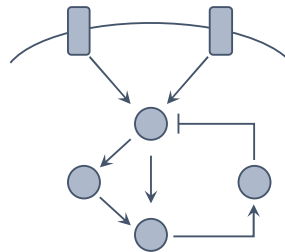
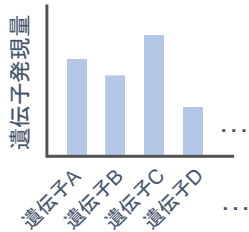


がん患者

遺伝子発現データ

数理モデル

患者固有シミュレーション



>>

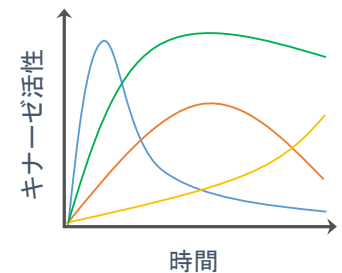


図 1 Pasmopy が可能にする患者固有モデリングの概要。生化学反応の説明文から数式を介さずに数理モデルが自動構築され、複数の細胞株データによってモデル内のパラメータを決定する。そこにがん患者由来の遺伝子発現データを入力することで、患者固有の薬剤応答などのシミュレーションが可能となる。

❖ 研究の背景および内容

生命科学分野では、細胞や組織の遺伝子発現や変異情報などのさまざまなデータが公共データベースに蓄積され、基礎・応用研究に利用されるようになりました。特に創薬研究においては、コスト、スピード、倫理面からも、このようなデータの利活用が重要視されています。解析手法としては、統計や機械学習などの手法が広く知られていますが、これらの手法は基本的に莫大な数の均一なデータを必要とし、相関などに基づいた薬剤や遺伝子の分類には適しますが、比較的少数のデータの解析や分子メカニズムなどの因果関係の予測は不得手なことから、新たな計算手法の開発が必要でした。

また、がん研究においては、従来のサブタイプ分類^{※8}を超えた患者の遺伝子情報に基づく層別化や個別化治療へのニーズが近年高まっています。しかし、研究現場では培養細胞やモデル動物を用いた細胞や組織内の分子活性を指標とする実験主導の研究が中心で、ヒト研究への橋渡しには、多くのコストとステップを必要としました。そのため、遺伝子情報を用い、コンピュータ上でひとりひとりの患者をバーチャルに再構成する“患者固有モデル”に関する研究が世界中で進められるようになりました。

しかし、遺伝子情報はあくまでもそのデータが取得された細胞や組織の一瞬の状態を捉えたもので、細胞などが置かれた生体環境における動的な振る舞いを反映していません。よって、このような患者固有モデル構築のためには、遺伝子情報に加え、患者の細胞本来の振る舞いを再現するための動的なパラメータが必要となります。そして、このパラメータの取得を行うためのパラメータ最適化には、本来、患者組織そのものから取得した細胞を培養し、多様な生育環境下で実験データを取得することが必要であり、この労力および技術的な困難さが患者固有モデルの開発のボトルネックとなっていました。また、患者固有の数理モデリングは、遺伝子発現データの入力から各患者の予後の予測や制御メカニズムを理解する上で強力なツールですが、モデルの構築や実装のためにはプログラミング技術やアルゴリズムなど、生命科学以外の専門的な知識や経験を必要とし、生物学や医学に関わる実験研究者が積極的に導入することを困難にしていました。

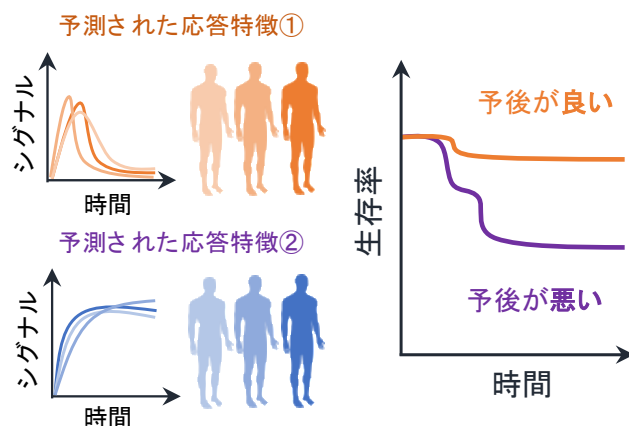
Press Release

そこで、本研究では、臨床の遺伝子発現情報から患者固有モデルを構築し、新たながん患者の層別化と潜在的な薬剤標的を患者レベルで予測することを目指しました。そして、複数の乳がん細胞株の遺伝子発現情報と実験データで学習したパラメータを数理モデルに導入し、がん患者の遺伝子発現データを入力することで患者固有モデルを実行できる枠組みを開発しました。その結果、患者固有のシミュレーション結果に基づくトリプルネガティブ患者の分類は、これまでの静的な遺伝子発現情報のみに基づく分類と比較して、より明確に予後の良し悪しを予測できることが明らかになりました。この結果は、患者固有モデルから得られる動的な情報が予後を予測するための新しいバイオマーカーとして利用できることを示唆しています(図2・左)。

本研究で構築された数理モデルは、これまでの文献情報に基づいた実際の反応系を記述しており、単なるがん患者の分類にとどまらず、制御メカニズムを患者ごとに解析することも可能にします。各患者のモデルを解析した結果、予後の悪いトリプルネガティブの患者群においてEGFR阻害剤への感受性がより低いことが示唆され、がん細胞株の薬剤応答データを解析することで、この予測の妥当性を確認しました。このように、イン・シリコ患者固有モデルは予後の予測のみならず、効果的な薬剤の探索にも使用できることを発見しました(図2・右)。

さらに、本研究では数理モデルの構築を飛躍的に容易にするために、対象となる生化学反応過程の文章を数式へと自動変換し、実行可能な数理モデルを作成する新しいモデル構築手法である「Text2Model」を開発しました。この手法を用いることで、数式を経由することなく、またそれをプログラムすることなく数理モデルをダイレクトに構築することが可能になります。さらに、説明文の可読性が極めて高いことから、より幅広い生命科学・医療分野のコミュニティで数理モデルの共有・編集・利用が促進されることが期待されます。

動的特徴量に基づく乳がん患者の再分類



潜在的な薬剤標的の予測

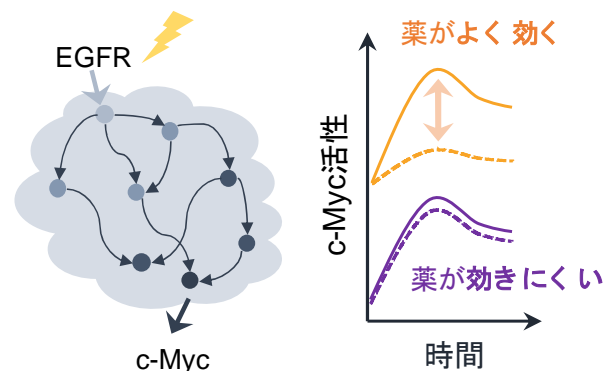


図2 (左) イン・シリコ患者固有モデルから得られる動的特徴量は従来の遺伝子発現情報のみに基づく分類よりもより明確にトリプルネガティブ患者の予後を分類する。(右) 患者固有モデルは予後の予測にとどまらず、潜在的な薬剤標的の予測にも利用可能。c-Mycは細胞のがん化に関連する転写因子。

❖ 本研究成果が社会に与える影響 (本研究成果の意義)

本研究により、公共データベースなどから患者の臨床検体から取得したRNAシーケンスデータが得られれば、実験をせずとも、直接、細胞シミュレーションにより、バーチャルに分子標的薬の効果を評価できる可能性が示されました。がん領域ではシグナル伝達系を構成するキナーゼを対象とした分子標的薬が多く開発され、その有効性の高さから、現在も活発に開発が進められています。これまでの創薬研究では、樹立培養細胞や患者から取得したオルガノイドを用いた実験解析が主体でしたが、樹立培養細胞では病態を正しく反映しておらず、オルガノイドに関してはその樹立に高度な技術を必要とすることが課題でした。今回開発した計算手法並びにシミュレーションとヒトの疾患ゲノム情報を組み合わせた手法を、遺伝子変異やエピゲノム情報などを含め、高度に発展することができれば、実験を行わなくとも、さまざまな公共データを組み合わせることで研究者の目的に合わせた解析が可能となります。また、細胞の数理モデルそのものもデータベースに多数登録されていますが、モデルが存在しないシグナル伝達系においても本解析ツールを用いて、数理科学の背景の無い生命科学研究者でも、テキストから容易に数理モデルを構築し、イン・シリコ患者の解析が

Press Release

可能となります。本研究の発展は、創薬研究のスピードアップ、コストの削減、患者ごとに適した薬剤を選択する個別化医療において大きなブレークスルーとなることが期待されます。

❖ 特記事項

本研究成果は、2022年3月11日(金)(日本時間)に国際科学誌『iScience』(オンライン)に掲載されます。

論文タイトル:

“A text-based computational framework for patient-specific modeling for classification of cancers”

著者:

Hiroaki Imoto, Sawa Yamashiro, Mariko Okada

掲載雑誌:

iScience (Cell Press)

DOI: 10.1016/j.isci.2022.103944

なお、本研究は、科学技術振興機構(JST)未来社会創造事業 探索加速型「共通基盤」領域の研究開発課題「創薬を加速する細胞モデリング基盤の構築(研究開発代表者:岡田 眞里子)」(JPMJMI19G7)、CREST バイオ DX 領域の研究開発課題「自然言語処理とシミュレーションによる細胞制御探索法の構築(研究代表者:岡田 眞里子)」(JPMJCR21N3)、および 日本学術振興会科学研究費補助金 基盤研究(A)「疾病機序理解のための遺伝子ネットワーク数理モデル基盤の構築」などの支援を得て行われました。

❖ 用語説明

※1. イン・シリコ

イン・シリコとは、生物学研究における試験管内(イン・ビトロ)および生体内(イン・ビボ)の実験条件に対応した造語で、計算機内の仮想条件を指します。

※2. 患者固有モデル(patient-specific model)

これまでのがん研究により、さまざまながんはひとつまたはそれ以上の遺伝子の変異の組み合わせにより起こること、またそのことは患者ごとに異なることが明らかになりました。そのため、これまでのような平均化した患者像でなく、コンピュータ上でひとりひとりの患者を、遺伝子情報をもとにバーチャルに再構成するイン・シリコの“患者固有モデル”が求められるようになりました。このような研究は海外で活発に進められており、患者の層別化や薬剤のオーダーメイド化に貢献すると考えられています。

※3. トリプルネガティブ乳がん

乳がんの治療標的となる3種類の受容体が欠如しており、難治性であり、乳がん全体の約 20%を占めます。他のタイプの乳がんと比較して、予後が不良であることが知られています。

※4. 公共データベース

生命科学研究では、培養細胞、モデル動物、ヒト疾患における遺伝子やたんぱく質の網羅的計測値は、データベースに保存されており、公共利用が可能です。例えば、米国 NIH の運営する The Cancer Genome Atlas(TCGA)プログラムでは、乳がんだけに限っても約 1,000 人の患者のゲノムおよび予後の情報が入手可能です。本研究では Cancer Cell Line Encyclopedia (CCLE)という培養細胞株のデータベースの乳がん細胞由来の遺伝子発現情報および TCGA がん患者の遺伝子発現データを利用し、解析に用いました。

Press Release

※5. 細胞シミュレーション

細胞シミュレーションに関する研究は、ヒトの全ゲノム解読前後の 2000 年頃から本格的に始まり、アメリカ、ドイツ、日本で活発に進められてきました。細胞の中の遺伝子間の相互作用や活性を数理モデルで記述し、コンピュータ上で再現する手法で、細胞の制御原理の理解や創薬に利用されています。細胞シミュレーションに用いられる数理モデルには、統計や物理法則など目的に応じて異なる原理を用いますが、本研究では、生化学反応の表現に用いられる連立微分方程式モデルを用いています。遺伝子の量や相互作用を情報として含むことから、疾患ゲノム解析により入手可能な遺伝子発現量や変異の情報などを入力値としてシミュレーションすることが可能になります。

※6. 感度解析

感度解析は、遺伝子発現量やパラメータの変化などの摂動を与えた場合の細胞ネットワーク出力に対する影響度(感度)を調べる手法です。本研究では、トリプルネガティブの乳がん患者固有モデルにおいて予後の良い患者群と悪い患者群で EGFR の感度係数に差異が見られ、そこから EGFR 阻害剤の感受性に違いがあると予測されました。この予測については、細胞株の薬剤応答データによってその妥当性が確認されました。

※7. EGFR 阻害剤

上皮成長因子受容体(EGFR)のチロシンキナーゼ部位を特異的に阻害して、EGFR から細胞内へがん細胞が増殖するための信号が伝わることを遮断することで、がんが大きくなるのを抑える、または、がんを小さくする低分子薬剤で、肺がんをはじめとしたがんの治療に広く使われています。

※8. サブタイプ分類

乳がんの場合、サブタイプは分子マーカーの発現の有無により、4つに分類されます。具体的には、エストロゲン受容体(ER)、ErbB2(HER2)受容体のどちらか一つが発現有、2つ共に発現有、2つ共に発現無という分類です。ERにはER拮抗剤など、ErbB2には抗体などの代表的薬剤があり、サブタイプに基づいた薬剤治療が現在行われています。しかし、最近のゲノム解析により、この4つのサブタイプ群それぞれにおいても、個々の患者の遺伝子の変異や発現の多様性が高く、効果的な薬剤の選択や治療のためには、新たな患者の層別化が必要だと考えられています。

❖ 本件に関する問い合わせ先

＜研究に関すること＞

大阪大学 蛋白質研究所 細胞システム研究室 教授
(兼:国立研究開発法人 医薬基盤・健康・栄養研究所 (NIBIOHN) 細胞ネットワーク制御プロジェクト・プロジェクトリーダー)

岡田 眞里子(おかだ まりこ)

〒565-0871 大阪府吹田市山田丘 3-2

TEL:06-6879-8617 FAX:06-6879-8619

E-mail: mokada[at]protein.osaka-u.ac.jp

＜JST事業に関すること＞

科学技術振興機構 未来創造研究開発推進部

庄司 真理子(しょうじ まりこ)

〒102-0076 東京都千代田区五番町 7 K's 五番町

TEL:03-6272-4004 FAX:03-6268-9412

E-mail: kaikaku_mirai[at]jst.go.jp



大阪大学
OSAKA UNIVERSITY

国立大学法人 大阪大学

〒565-0871 大阪府吹田市山田丘 1-1

TEL: 06-6877-5111 (代)

www.osaka-u.ac.jp

Press Release

<報道担当>

大阪大学 蛋白質研究所 広報室

E-mail: [kouhou\[at\]protein.osaka-u.ac.jp](mailto:kouhou[at]protein.osaka-u.ac.jp)

科学技術振興機構 広報課

〒102-8666 東京都千代田区四番町 5 番地 3

TEL:03-5214-8404 FAX:03-5214-8432

E-mail: [jstkoho\[at\]jst.go.jp](mailto:jstkoho[at]jst.go.jp)