

## 科学研究費助成事業 研究成果報告書

平成 29 年 6 月 12 日現在

機関番号：14301  
 研究種目：若手研究(B)  
 研究期間：2013～2016  
 課題番号：25870336  
 研究課題名(和文) Drug candidate discovery by development of a context-sensitive target network similarity metric  
 研究課題名(英文) Drug candidate discovery by development of a context-sensitive target network similarity metric  
 研究代表者  
 Brown John (Brown, John)  
 京都大学・医学研究科・講師  
 研究者番号：90583188  
 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)：本研究では、計算創薬の予測技術研究を行った。多くの薬の原理は、化合物をタンパク質に結合させ、生体内の信号伝達を阻止することである。どの化合物とどのタンパク質が結合するかは複雑であるが、最近では計算で相互作用を予測することができ、薬の開発費用を抑えることが期待されてきた。この研究では相互作用を予測する、新しい予測技術を開発した。新手法の特徴は、膨大なデータまたは複雑な人工知能(AI)を用いずに、正確な予測モデルを構築することができること。本研究では主流になってきたビッグデータ創薬を覆す手法であり、手法の再現性も証明できた。今後は、製薬会社と連携して実際の医薬品開発に適用する。

研究成果の概要(英文)：In this project, we executed research for computational drug discovery. The fundamental principle of most drugs is that a compound works to inhibit the function of a particular protein. To know in advance which compounds will inhibit which proteins is difficult, but in recent years computational methods have become developed which are expected to bring down drug development costs. In this research we developed a new method for the prediction of compound-protein interactions. Among its many features, ours is special in that it requires neither big data nor complex artificial intelligence (AI), yet can still build a highly predictive model. We disproved the current trend of "big data drug discovery" by developing a reproducible method and reported multiple scientific papers. We are planning with pharmaceutical companies to apply the technology to joint research projects.

研究分野：計算創薬

キーワード：chemogenomics comput. drug discovery pattern recognition life science informatics ケモジエノミクス 計算創薬 統計パターン認識 生命情報科学

### 1. 研究開始当初の背景

病気を治療する多くの場合に、病気の原因となると考えられるタンパク質に小型化合物を与えることでそのタンパク質を阻止する作戦がある。創薬研究はどの化合物が標的タンパク質に結合して薬効を及ぼすかを調べる分野である。化合物は実質的に無数であるため、全ての化合物を各タンパク質に合わせて活性を測ることが不可能である。また、10万や100万化合物だけの薬効を測ろうとしても、膨大な費用がかかる問題がある。さらに、タンパク質に結合して薬効を示す化合物の割合は2~3%に過ぎないため、多くの実験は無駄になる。

この問題を克服するために創薬研究分野ではコンピューターによる活性予測が近年に多くの注目が集まっている。「仮想スクリーニング」という研究手法は、化合物の化学構造をコンピューターに入力するだけで、その化合物の活性(薬効や毒性など)を予測して、薬効を示すと予測されない化合物を排除して候補化合物を効率的に同定する手法である。

予測することは、「予測モデル」を構築することに等しい。予測モデルは、多くの場合に数式であるが、数式以外の方法もある。例えば、予測対象となる化合物をいくつかの決まった基準と順次に比較して、その化合物が全ての基準を満たせば薬効があると判断する、という基準も考えられる。後者の方法は、「決定木」という。

一方、どの活性データを用いて予測モデルを作成するのが未確定である。例えば、一つのタンパク質に注目して少量の活性データからモデルを作成する方法がある。また、標的タンパク質のX線結晶構造を用いて物理化学の方式に従って立体的シミュレーションの結果から薬効を評価する方法もある。また、近年注目されているもう一つの方法は、化合物とタンパク質の膨大な相互作用情報から相互作用(活性)する組み合わせとしない組み合わせ(不活性)を識別する手法もある。この最後の方法は、「ケモジェノミクス」という手法で、本研究に用いた方法です。

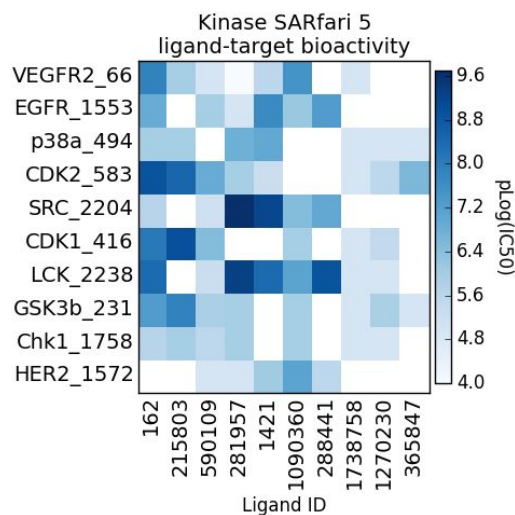
上記に述べたが、創薬研究の中心的な問題は、どの化合物がタンパク質を阻害するか、とのことであり、それを調べるには膨大な費用がかかる。そこで、計算創薬では、活性データから予測モデルを算出して、効率よく活性を示す化合物を同定することを目標とする。

### 2. 研究の目的

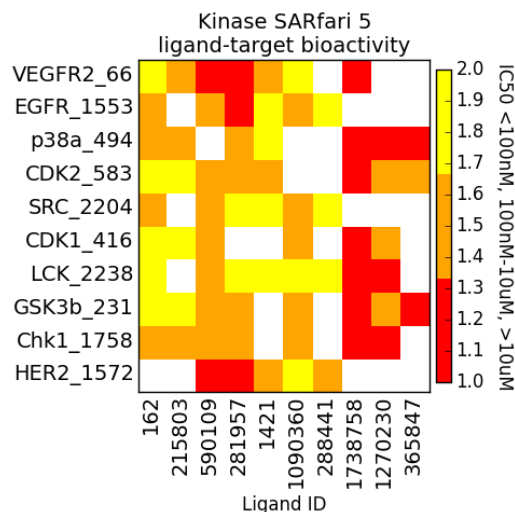
本研究では「ケモジェノミクス形」の高予測能力を持つ計算創薬予測モデルをどのように作成できるかを研究してきた。2014年以降に注目されている「ビッグデータ」計算予測モデルの有用性が十分示されていないところ、本研究では逆方向で「スモールデータ」の高予測能力モデル構築法に挑んだ。

### 3. 研究の方法

まず、化合物とタンパク質の活性データを収集して品質管理を行った。欧州立分子生物学研究所(EMBL)のホームページには、有名な創薬標的タンパク質ファミリーの活性データが公開されている。例えば、下記の図では化合物の酵素(キナーゼ)への阻害活性を示すデータがEMBLの公開データにある。色の意味は、活性の強さを示すこと。



次に、このデータを「活性有」と「活性無」に変換する。その理由は、化合物とタンパク質の活性を定量的に予測することが困難であるためである。また、実験では計測にバラツキがあるため、一つの活性値を厳密に予測する意味がないと考えられる。従って、上記のようなデータを「強く活性する」、「強く活性しない」、そして「中間的な活性を示す」3種類に分ける。この場合、創薬研究で多く適用される活性値を基準にして元データを離散化させる。上記のデータを離散化させた結果は下記の通りである：



上記のデータから、「強く結合する」(黄色)組み合わせと「強く結合しない」(赤色)組み合わせを残して、オレンジ色を排除する。

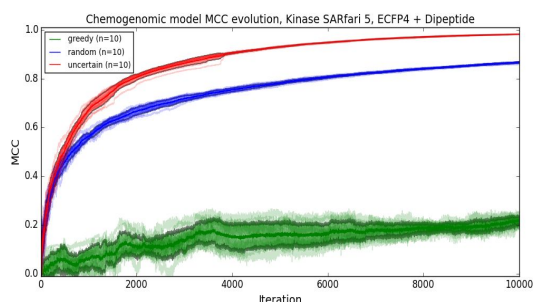
次に、「強く結合する」と「強く結合しない」のデータをモデルに入力する方法を研究した。「ビッグデータ」計算創薬では、大量な相互作用データを一気に人工知能に提示して、一つの予測モデルをもとめる方法がある。しかし、この研究では2013年に「ビッグデータ」計算創薬モデルを用いて国際共同研究で予測実験を評価した結果、予測精度は低かった。「低い」という評価に用いた尺度は、マッシュズ相関係数という統計であり、単純な「正解の数」という統計よりも厳密に予測精度を評価する方法である。

「ビッグデータ」に基づく予測実験に失敗した後、スイス連邦大学チューリッヒ校の共同研究者と一緒に新計算創薬手法の開発を開始した。その方法は、膨大なデータを一つのモデル計算プログラムに入力するのではなく、**活性の予測に最も必要なデータのみを用いて簡単な予測モデルを構築する**、計算創薬の主流と全く異なる戦略をもつ方法である。本研究では、必要なデータを選択する方法を二つ開発してきた。一つ目は、「貪欲法」(英: greedy)法である。この方法では、活性を示す(上記の図では黄色点)を優先的に選択する。もう一つは、「好奇心法」(英: curiosity)法である。この方法では、予測に一番判断しにくいデータを優先的に選択する方法である。

また、二つの手法との比較のため、活性点をランダムに選択する方法と比較を行った。

#### 4. 研究成果

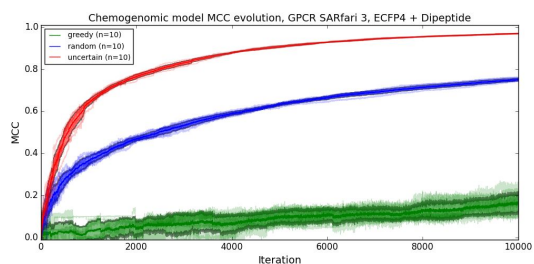
酵素(キナーゼ)の阻害活性をもつ化合物の活性を予測する実験結果は以下の図にある。横軸は、予測モデルに含まれるデータの数である。縦軸は、マッシュズ相関係数(すなわち、予測精度)を示す。「貪欲法」の予測精度は緑色で示されている。「好奇心法」は赤色で示されている。結果の再現性を評価するために、各実験を10回ずつ行った後、平均予測精度と分散を可視化した。下記の図では、不活性を含める元活性データの数は、約5満点がある。



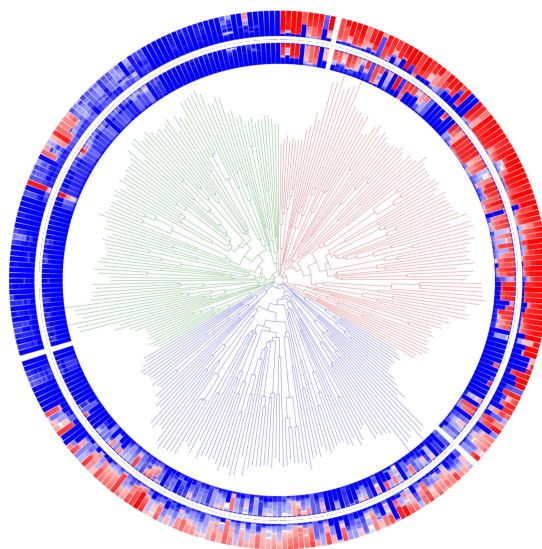
上記の図から、いくつかの結果が明らかになった。一つは、ランダムに化合物とタンパク質を選択するよりも共同研究で設計してきた「好奇心法」が優れていることである。二つ目は、膨大なデータがあっても、高予測精度を示すモデルに**全てのデータを入力する必要はない**。

酵素(キナーゼ)以外に、2種類の細胞膜に存在するタンパク質(GPCR)データベースも用いて提案手法を評価した。GPCRは製薬会社では非常に注目されているタンパク質の種類であり、現在市販されている医薬品の多くがこのGPCRを標的にしている。ドパミンやセロトニンなどのホルモンの受容体は、GPCRである。

GPCR SARfari というデータベースの予測実験結果は下記の図の通り:



開発してきた手法の有用性を検証するもう一つの方法は、各タンパク質に対する予測精度を測る方法である。ここで、タンパク質間の類似度を用いて系統樹を描画し、系統樹の葉に、各タンパク質の予測精度をヒートマップとして付け足した。この可視化法を用いることにより、タンパク質ファミリー全体の予測性を一目で評価することができる。例えば、GPCRファミリーにおける各タンパク質の予測性は、次の図に示されている:



ここでは、内側の系統樹の色は左図の色と同様な意味を示す。本研究で開発した方法は、赤と緑の系統樹で示されており、外側のヒートマップは一つずつのタンパク質に対する予測能力を示している。上記の図で明らかになっているのは、本研究で開発した「好奇心法」の予測能力はランダム選択法より優れていること。従って、本研究で開発してきた手法は、製薬会社の開発で利用できると考えられる。

5. 主な発表論文等

〔雑誌論文〕(計 4 件)

(2013) **J.B. Brown**, Satoshi Nijjima, Yasushi

Okuno

Compound-protein interaction prediction within chemogenomics: theoretical concepts, practical usage, and future directions

*Molecular Informatics*, 32:906

DOI: 10.1002/minf.201300101 -- Pubmed ID 27481137

(2014) **J.B. Brown**, Yasushi Okuno, Gilles

Marcou, Alexandre Varnek, Dragos Horvath

Computational chemogenomics - is it more than inductive transfer?

*Journal of Computer-Aided Molecular Design*, 28:597

DOI: 10.1007/s10822-014-9743-1 -- Pubmed ID 24771144

(2014) **J.B. Brown**, Masahiko Nakatsui, Yasushi

Okuno

Constructing a foundational platform driven by Japan's K supercomputer for next-generation drug discovery

*Molecular Informatics*, 33:732

DOI: 10.1002/minf.201400067 -- Pubmed ID 27485419

\*\*(2017) Daniel Reker, Petra Schneider, Gisbert

Schneider, **J.B. Brown**

Active learning for computational chemogenomics

*Future Medicinal Chemistry*, 9(4):381-402

DOI: 10.4155/fmc-2016-0197 -- PubMed ID 28263088

この論文は、日本経済新聞電子版で報道され、国内外に注目が集まっている：  
( [http://www.nikkei.com/article/DGXLRS438913\\_Y7A300C1000000/](http://www.nikkei.com/article/DGXLRS438913_Y7A300C1000000/) )

〔学会発表〕(計 3 件)

Do you really need big data for drug discovery and repurposing? A critical investigation.

*Tokyo Medical and Dental University 15th Surugadai International Symposium : Current Status and Future of Drug Discovery*, Tokyo, Japan, 2016

Computational chemogenomics: core concepts, applications, and critical aspects of model building

*Tokyo Denki University Special Lectures on Recent advancements in chemoinformatics and structural informatics*, Tokyo, Japan, 2016

Proteochrometric virtual screening: a launchpad to small molecule discovery and design

*Institute for Transformative Bio-Molecules, Small Molecule Discovery Colloquium*, Nagoya University, Nagoya, Japan, 2015

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況 (計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

ブラウン・ジョン (John BROWN )  
京都大学大学院・医学研究科 講師  
研究者番号：90583188

### (2) 研究分担者

### (3) 連携研究者

### (4) 研究協力者

シュナイダー・ギズバート  
(Gisbert SCHNEIDER )  
ETH Zurich, Switzerland

レカー・ダニエル  
(Daniel REKER)  
Massachusetts Institute of Technology,  
USA