

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 20 日現在

機関番号：84420

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26870045

研究課題名(和文) 配列情報に基づく蛋白質相互作用予測の高度化と創薬標的探索のための支援システム開発

研究課題名(英文) Development of a sequence-based prediction method of interacting protein partners for drug target discovery

研究代表者

村上 洋一 (Murakami, Yoichi)

国立研究開発法人医薬基盤・健康・栄養研究所・医薬基盤研究所 創薬デザイン研究センター・プロジェクト研究員

研究者番号：20548424

交付決定額(研究期間全体)：(直接経費) 1,600,000円

研究成果の概要(和文)：本研究では、研究代表者が開発したタンパク質間の相互作用の可否を予測するPSOPIAを応用して、不均衡かつ大規模な訓練データセットを用いて、ある1つのタンパク質と相互作用をする他のタンパク質を予測する方法を開発した。新たに開発した機械学習モデルは、以前のPSOPIAよりもより高い予測性能を達成することができた。また、ある1つのタンパク質と公共データベースに登録されているヒトの各タンパク質とをペアにして自動的かつ高速に相互作用予測を実行するシステムを構築した。さらに、予測されたタンパク質から創薬標的やバイオマーカーとなりうるタンパク質の絞り込みを支援するためのインターフェースを開発した。

研究成果の概要(英文)：This study applied previously developed PSOPIA, a method for predicting an interaction between two proteins, for developing a new computational method to predict proteins that interact with a given protein. The new method was trained on a highly imbalanced and large training dataset of human protein-protein interactions and successfully achieved higher prediction performance than PSOPIA. I have also developed a fast system that can automatically make predictions of interactions between a given protein and all the proteins registered in a public database. Furthermore, I have developed a graphical user interface with the new method developed in this study for identifying drug targets and biomarkers.

研究分野：バイオインフォマティクス

キーワード：タンパク質間相互作用 タンパク質機能 機械学習 創薬支援

### 1. 研究開始当初の背景

タンパク質間の相互作用は、多くの生物学的に重要な機能を成立させている。従って、ある特定のタンパク質と相互作用する他のタンパク質を同定し、それらの相互作用ネットワークや相互作用メカニズムを解明することは、その特定のタンパク質の生物学的な機能の理解だけでなく疾患に関わるパスウェイ等を理解する上でも非常に重要である。

そのようなタンパク質間の相互作用の重要性から、様々な相互作用が実験的に同定されている。しかしながら、その作業は手間や費用がかかるだけでなく、限られた相互作用しか同定することができない等の制約がある。そのため、例えば、疾患との関与が疑われるタンパク質が発見されたとしても、他のタンパク質との相互作用に関する情報が不明なことが多く、その疾患関連タンパク質の生物学的機能を理解することが困難な場合がある。

そこで研究代表者は、機械学習法を用いて、タンパク質のアミノ酸配列情報のみから他のタンパク質と相互作用する部位(アミノ酸残基群)を予測する方法(<http://mizuguchi lab.org/PSIVER>;[1])を開発した。実際にこの方法を用いて、乳がん細胞中に存在するタンパク質に対して、がん細胞の増殖に関与する部位を予測し、その部位に特異的に結合する他のタンパク質との結合を阻害する分子設計をすることによって乳がん細胞の増殖を抑えることに成功した[2]。

また研究代表者は、2つのタンパク質のアミノ酸配列が与えられたときに、それらが相互作用する可能性が高いのかを予測する方法を開発しウェブ上に公開している(<http://mizuguchi lab.org/PSOPIA>;[3])。しかしながら、この方法は2つの配列の入力が必要であるという制約や、予測した結果の解釈を利用者に頼っているのが現状であり、実際問題として、機能が全く未知のタンパク質への適用が難しいという問題がある。

このような問題を解決するための新たな予測方法を開発することは、多くのタンパク質の生物学的機能の解明、また疾患関連タンパク質とその周辺タンパク質から創薬標的を効果的に探索することに大いに貢献すると考えられる。

### 2. 研究の目的

本研究では、ある特定のタンパク質と相互作用する他のタンパク質を予測し、そして予測されたタンパク質から創薬標的やバイオマーカーになりうるタンパク質の絞り込みができるような支援システムの開発を目的として、次の目標を設定した。

(1) 与えられた1つのタンパク質と相互作

用する他のタンパク質を予測する方法の開発

(2) タンパク質の相互作用部位を予測する方法の高度化

(3) (1)で開発した方法が出力する予測結果に、(2)を含めた各種アノテーションを統合したグラフィカルな出力インターフェースの開発

### 3. 研究の方法

(1) 研究代表者がこれまでに開発した、2つのタンパク質のアミノ酸配列のみからそれらの相互作用の可否を予測する方法(PSOPIA; [3])を、与えられた1つのタンパク質のアミノ酸配列のみからそれと相互作用する他のタンパク質を予測する方法に応用する。すなわち、与えられた1つのアミノ酸配列と、公共的に利用可能なタンパク質配列データベースから得られた1つのアミノ酸配列とをペアにして、それらの相互作用を予測する。そして、この予測を公共データベースに登録されているヒトのアミノ酸配列全てに対して繰り返し実行する計算システムを開発する。また、PSOPIAの開発以降に実験的に確かめられた相互作用データを加えた不均衡かつ大規模な訓練データセットを準備して、機械学習法のパラメータ値を最適化し予測性能の向上を目指す。

(2) 研究代表者がこれまでに開発した、タンパク質のアミノ酸配列情報のみから、他のタンパク質と相互作用する可能性が高いアミノ酸残基(以下、相互作用部位と呼ぶ)を予測する手法(PSIVER)は、各残基位置における特徴量が独立していると仮定するベイズの定理に基づいた確率分類器である。この独立性の仮定を緩和した新たな機会学習法を適用することによって予測方法の高度化を目指し、必要なベンチマークを実行する。

(3) (1)で開発した方法を用いて、与えられたタンパク質と相互作用する可能性があるとして予測されたタンパク質の予測スコアに基づいて順位付けをする出力インターフェースを開発する。また、予測された複数のタンパク質の中から、さらに可能性が高いタンパク質を絞り込むために必要な各種公共データベースへのリンクやパスウェイやGO解析結果等の統合した出力インターフェースを検討する。さらに、与えられたタンパク質とその周辺タンパク質の既知の相互作用ネットワーク上に、予測された相互作用の情報を加えた相互作用ネットワーク図を表示できる出力インターフェースも開発する。

### 4. 研究成果

(1) ゲノムやパスウェイ情報、またタンパク質の構造や機能情報から化合物情報までを幅広く統合しているTargetMine データウェアハウス(<http://targetmine.mizuguchilab>).

org; [4]) を利用して、そこに格納されている実験的に確認されたヒトのタンパク質間相互作用データのうち高信頼度（論文の引用件数が2件以上）のものを取得し、それらを利用して正例と負例の数が極端に偏った不均衡かつ大規模な訓練データセットを作成した。そして、特徴間の緩い依存関係を考慮した Averaged One-Dependence Estimator (AODE) を用いて、新たに機械学習モデルを構築しベンチマークを実行した。その結果、予測性能の向上に成功した（図1；以前の性能：AUC=0.79, pAUC<sub>0.5%</sub>=0.16 [3]）。

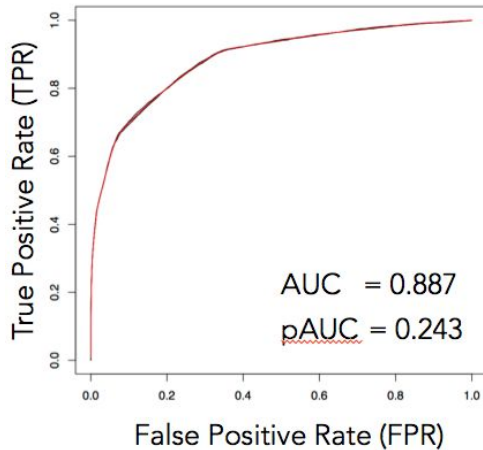


図1. 新たに開発した AODE モデルの予測性能

この新たな機械学習モデルを利用して、与えられた1つのタンパク質と公共データベースの UniProtKB に登録されているヒトの各タンパク質とをペアにして自動的にかつ高速に相互作用予測を実行するシステムを開発した（以下、PSOPIA-X と呼ぶ；論文準備中）。

(2) タンパク質の相互作用部位や非相互作用部位に含まれる各アミノ酸残基の特徴が独立していると仮定するベイズの定理に基づいた確率分類器ではなく、その仮定を緩和した Averaged One-Dependence Estimator (AODE) を用いた機械学習モデルの構築を行い、ベンチマークを実行した。従来の PSIVER の予測性能と比較するために、それを開発した際に使用した同様の訓練データセットを用いて、新たに AODE モデルを作成した。その結果、特徴間の依存関係を考慮しない従来法よりも高い性能を得ることができた。しかしながら、従来法で予測された真陽性が新手法では偽陰性になり、またその逆もあることがベンチマーク結果の解析から明らかになった。これらの方法の併用という選択肢もあるが、将来的に両方の予測スコアを特徴量として機械学習することでより信頼のある予測ができるメタアプローチの開発の必要性に至った。

(3) (1)で開発した機械学習モデルを利用して、相互作用する可能性が高いと予測された

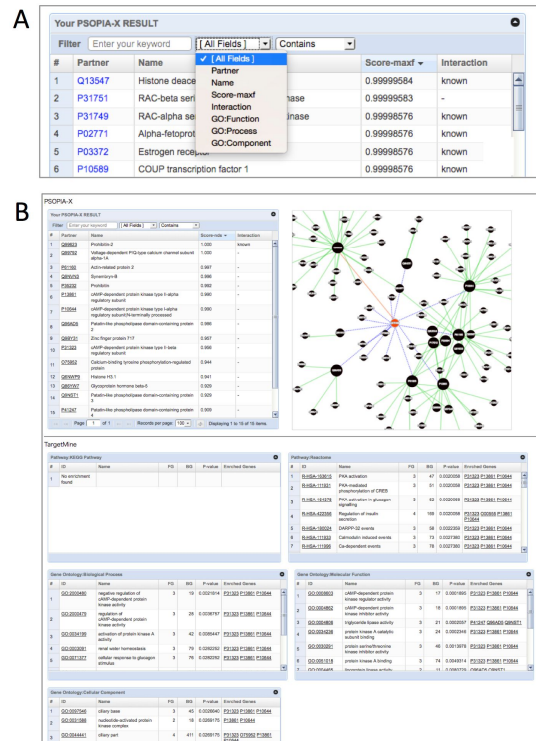


図2. 創薬標的の絞り込みに役立つ各種アノテーションを統合した PSOPIA-X のインターフェース

タンパク質を予測スコアに基づいて順位付けしたリストを表示するインターフェースを開発した（図2）。これは JavaScript 言語の JQuery ライブラリーを用いたグラフィカルなインターフェースであり、並べ替えや検索をすることができる。予測されたタンパク質は、予測スコア順に並べられ、タンパク質名や与えられたタンパク質との相互作用が既知かどうかの情報、また遺伝子オントロジーによる機能や細胞内局在等の情報を得ることができる（図2-A）。また各タンパク質のより詳細なアノテーションを得るために UniProtKB へのリンクが貼られている。予測されたタンパク質リストから TargetMine を用いてパスウェイや遺伝子オントロジーの解析した結果も表示される（図2-B）。また、与えられたタンパク質とその周辺タンパク質の既知の相互作用ネットワーク上に、予測された相互作用を加えて相互作用ネットワークを Cytoscape プラグインによりインターフェース上に表示した。このような機能により利用者が興味のあるタンパク質の周辺タンパク質との関係性を可視化しながら創薬標的の絞り込みができるようになった。さらに、予測したタンパク質リストから創薬標的になりうるタンパク質の絞り込みを支援するために、ゲノムやタンパク質の構造や機能の情報、またパスウェイ情報や化合物情報、ゲノムワイド関連解析 (GWAS) から得られた疾患に関与する一塩基多型 (SNP) や相互作用情報までを幅広く統合している TargetMine と PSOPIA-X との統合を行った。

本研究で実施したタンパク質間相互作用予測の高度化と創薬標的探索のための支援システムの開発を通じて調査した、研究代表者が開発した予測方法を含めた各種タンパク質間相互作用の予測方法や関連技術等を整理し、海外の論文誌に総説論文として発表した (Murakami *et al.*, *Current Opinion of Structural Biology*, 2017)。

#### <引用文献>

- [1] Murakami, Y. and Mizuguchi, K., (2010) *Bioinformatics*, 26, 1841-8.
- [2] Yoshimaru *et al.*, (2013) *Nature communication*, 4:2443.
- [3] Murakami, Y. and Mizuguchi, K., (2014) *BMC Bioinformatics*, 15:213.
- [4] Chen, Y.A., Tripathi, L.P., and Mizuguchi, K., (2011) *PLoS One*, 6, e17844.
- [5] Murakami, Y., Omori, S. and Kinoshita K., (2016), *Journal of Structural and Functional Genomics*, 17(4):101-110.

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

##### [雑誌論文](計4件)

Yoichi Murakami and Kenji Mizuguchi, Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators, *BMC bioinformatics*, 査読有, 15:213, 2014, DOI: 10.1186/1471-2105-15-213  
Yi.An Chen, Yoichi Murakami, Shandar Ahmad, Tetsuro Yoshimaru, Toyomasa Katagiri and Kenji Mizuguchi, Brefeldin A-inhibited guanine nucleotide-exchange protein 3 (BIG3) is predicted to interact with its partner through an ARM-type alpha helical structure, *BMC Research Notes*, 査読有, 7:435, 2014,

DOI:10.1186/1756-0500-7-435

Yoichi Murakami, Satoshi Omori and Kengo Kinoshita, NLDB: a database for 3D protein-ligand interactions in enzymatic reactions, *Journal of Structural and Functional Genomics*, 査読有, vol.17, issue.4, 2016, pp.101-110,

DOI:10.1007/s10969-016-9206-0

Yoichi Murakami, Lokesh P Tripathi, Philip Prathipati and Kenji Mizuguchi, Network analysis and in silico prediction of protein-protein interactions with applications in drug discovery, *Current Opinion of Structural Biology*, 査読有, vol.44, pp.134-142, 2017, DOI:10.1016/j.sbi.2017.02.005.

##### [学会発表](計6件)

村上洋一、大森聡、木下賢吾、Prediction of 3D protein-ligand interactions in enzymatic reactions、構造活性相関シンポジウム、2016.11.16、京都大学医学部創立百年記念施設・芝蘭会館(京都市)

村上洋一、水口賢司、Prediction of interacting proteins from homologous protein-protein interactions、CBI学会 2016年大会、2016.10.25-27、タワーホール船堀(東京都江戸川区)

村上洋一、水口賢司、Prediction of interacting protein pairs using homology-based approach、第5回生命医薬情報学連合大会、2016.9.29~10.1、東京国際交流館 プラザ平成(東京都江東区)

村上洋一、大森聡、木下賢吾、A database for protein-ligand complexes in enzymatic reaction、CBI学会、2015.10.27、タワーホール船堀(東京都江戸川区)

##### [図書](計0件)

##### [産業財産権]

##### 出願状況(計1件)

名称:「エピトープ均質化パネル、ならびにその作製方法および利用」

発明者:近藤裕郷、鎌田春彦、永田諭志、水口賢司、村上洋一

権利者:国立研究開発法人医薬基盤・健康・栄養研究所

種類:特許

番号:特願 2017-21553号

出願年月日:2017年2月8日

国内外の別:国内

##### 取得状況(計0件)

##### [その他]

##### ホームページ等

PSOPIA - アミノ酸配列情報に基づく相互作用タンパク質の予測サーバー

<http://mizuguchi lab.org/PSOPIA/>

TargetMine - 創薬の初期研究を支援するデータウェアハウス

<Http://targetmine.mizuguchi lab.org>

NLDB - 酵素反応に関わるタンパク質と天然化合物との複合体構造データベース

<http://nlldb.hgc.jp>

#### 6. 研究組織

(1)研究代表者

村上 洋一 (MURAKAMI, Yoichi)  
国立研究開発法人医薬基盤・栄養・健康研  
究所・創薬デザイン研究センター・プロジ  
ェクト研究員  
研究者番号：20548424

(2)研究分担者  
なし

(3)連携研究者  
なし

(4)協力研究者  
なし