

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 6 日現在

機関番号：17102

研究種目：若手研究(A)

研究期間：2013～2015

課題番号：25700029

研究課題名(和文)オミックスデータから薬物の標的分子を網羅的に予測するための機械学習法の開発

研究課題名(英文)Development of machine learning methods to comprehensively predict drug targets from omics data

研究代表者

山西 芳裕 (Yamanishi, Yoshihiro)

九州大学・高等研究院・准教授

研究者番号：60437267

交付決定額(研究期間全体)：(直接経費) 18,500,000円

研究成果の概要(和文)：薬物・標的タンパク質間相互作用の同定は、医薬品開発において最重要課題である。ポストゲノム研究では、ゲノム、トランスクリプトーム、プロテオームなどの遺伝子やタンパク質に関する大量のオミックスデータが得られるようになってきた。同時に、膨大な数の化合物や薬物に関するケミカル情報や生理活性情報も蓄積されている。本研究プロジェクトでは、そのような薬物やタンパク質に関する膨大なオミックスデータを融合解析し、未知の薬物・標的タンパク質間相互作用を予測するための機械学習の手法を開発することを目的としている。

研究成果の概要(英文)：The identification of interactions between drugs and target proteins is a key area in genomic drug discovery. Various high-throughput experimental projects enable us to analyze the genome, transcriptome, proteome. At the same time, the high-throughput screening of large-scale chemical compound libraries with various biological assays produce chemical and phenotypic data on drugs and bioactive compounds. In this project, we develop machine learning methods to infer unknown drug-target interactions by integrating various large-scale omics data.

研究分野：バイオインフォマティクス

キーワード：機械学習 インシリコ創薬 標的分子 相互作用予測 薬物

1. 研究開始当初の背景

大半の医薬品(薬物)は標的とするタンパク質などの生体分子との相互作用を介して、薬効や副作用という形で人体に影響を及ぼす。そのため、薬物・標的タンパク質間の相互作用の同定は、医薬品の開発過程において最重要課題である。ポストゲノム研究では、ゲノム、トランスクリプトーム、プロテオームなどの遺伝子やタンパク質に関する大量のオミックスデータが得られるようになってきた。同時にハイスループットスクリーニングなどの技術の発展によって、膨大な数の化合物や薬物に関するケミカル情報や生理活性情報も蓄積されている。そのため、これらの様々なオミックスデータを有効活用し、薬物・標的タンパク質間相互作用を大規模に予測するための情報技術の開発が期待されている。

このような状況の中、薬物・標的タンパク質間相互作用ネットワーク予測法の研究が近年盛んに行われている。これまでの先行研究では、“類似薬物は類似タンパク質と相互作用しやすい”という仮定のもと、薬物の化学構造類似性とタンパク質の配列類似性を用いる「ケミカルゲノミクス」と薬物の副作用類似性とタンパク質の配列類似性を用いる「薬理ゲノミクス」の2つの枠組みがある。いずれかの枠組みにおいて、薬物・標的タンパク質間相互作用を予測するための様々な手法が提案されてきている。

実際の創薬の現場においては、相互作用の有無の予測だけではなく、例えば、薬物の化学部分構造、薬効や副作用、タンパク質の機能モチーフ、パスウェイがどのように相互作用に関連しているのかを理解するための手がかりを得たいという要望が強い。そのため、薬物の標的タンパク質予測の精度を高めると共に、予測過程を説明できるような相互作用パターンを表す分子の特徴の抽出は挑戦的な課題である。

2. 研究の目的

本研究では、薬物やタンパク質に関する様々なオミックスデータから薬物・標的タンパク質間相互作用のパターンを表す分子の特徴を抽出し、未知の相互作用をゲノムワイドに予測するための機械学習の手法を開発する。先行研究の研究では、主にカーネル法という機械学習の手法に依存してきた。カーネル法は構造データを扱いやすいという長所はあるが、特徴抽出ができないため予測過程がブラックボックスになるという短所があり、相互作用メカニズムを生物学的に説明できる分子の特徴を利用者に提供できなかった。そこで本研究プロジェクトでは、薬物・標的タンパク質間相互作用パターンを表す特徴をオミックスデータから抽出し、それを基に未知の相互作用ネットワークを予測する手法を提案する。

3. 研究の方法

本研究で提案する手法を開発するところから始める。まず、各薬物・タンパク質ペアをフィンガープリント(高次元のバイナリ特徴ベクトル)で表すことを考える。化学部分構造や副作用などの薬物の情報と、機能モチーフやパスウェイなどのタンパク質の情報を融合し、異種データ間の相互作用を表せるようなフィンガープリントを構築する方法について検討する。グラフや文字列などの構造データをフィンガープリントに変換する様々なアルゴリズムを検討する。

次に、薬物・タンパク質間相互作用を予測する手法を開発する。具体的には、機械学習の分野で特徴抽出能力を持つ予測モデルとして有望視されているスパース統計手法を、薬物・タンパク質ペアに適用し、各ペアを相互作用するかしないかを予測する分類問題を考える。高次元のバイナリデータに特化した場合、大規模計算にも耐えうる高速なアルゴリズムを実装する。生物学的に意味のある特徴抽出を行うために本研究では特徴間の関連構造のモデル化を検討する。これによって薬物の化学部分構造の階層性、副作用の語彙の階層性、パスウェイの階層性などの複雑な構造を考慮した特徴抽出が可能になることが期待できる。

上記の手法の開発とともに、医薬品関連データの収集も同時に行う。薬物の標的分子やオフターゲットの情報は、DrugBank データベースや STITCH データベースから得る。薬物の副作用データは、米国 FDA(食品医薬局)で公開されている数百万人の患者に対する市販後調査報告書を、テキストマイニングして薬物と副作用の関係を得る。薬物の化学構造、タンパク質の配列情報、タンパク質の機能モチーフ情報、パスウェイ情報は、KEGG データベースから得る。

4. 研究成果

初年度(平成25年度)に、薬物やタンパク質に関連するデータの収集を行った。薬物の化学構造、標的タンパク質やオフターゲットの情報、薬物の副作用データ、タンパク質の機能ドメイン情報は、様々な文献や様々な公共データベース(KEGG, DrugBank, Matador, ChEMBL, BindingDB, TTD, PubChem, SIDER, FAERS, PFAM など)から得て、データを情報解析可能な形に整備した。次年度以降(平成26年度や平成27年度)も、適時必要なデータの収集や更新を行った。

それらのデータに基づき、薬物やタンパク質の記述子、薬物・タンパク質ペアを表すフィンガープリントを開発した。薬物の記述子とタンパク質の記述子を別々に構築し、それをテンソル積で繋ぎ合わせることによって、各薬物・タンパク質ペアをフィンガープリントで表した。またフィンガープリントに minwise hashing というアルゴリズムを適用

することで、情報をなるべく落とさずに低次元のフィンガープリントに変換する方法を開発した。

薬物・タンパク質ペアを表す方法として、高次元のバイナリ特徴ベクトルであるフィンガープリントだけでなく、高次元の連続値特徴ベクトルである記述子の構築を行った。薬物の化学構造を表現するために、官能基や原子の環境情報を考慮できる KCF-S (KEGG Chemical Function Substructure) という記述子を開発した。これによって薬物の部分構造の階層関係を考慮した記述子を実現した。薬物の薬理作用情報を表現するため、FDA AERS に登録されている約 1 万種類の副作用の報告頻度で薬理作用プロファイルを表現した。フェノタイプのオントロジーを用いることによって、薬物の薬理作用の階層関係を考慮した記述子を実現した。タンパク質に関しては、パスウェイ、連続アミノ酸の頻度や機能ドメインの頻度によるプロファイルを表現した。これらを組み合わせて薬物・タンパク質ペアの特徴を高次元の連続値ベクトルで表す記述子を構築した。

次に、薬物・タンパク質ペアのフィンガープリントや記述子を基に、ペアワイズなカーネル回帰やスパース分類器を適用し、薬物・タンパク質ペアの相互作用ネットワークを解析および予測した。スパース分類器では、L1 罰則化に基づくサポートベクターマシンや L1 罰則化に基づくロジスティック回帰などを適用し、各薬物・タンパク質ペアの相互作用を解析および予測した。同時に、相互作用に関連があると推測される薬物の化学部分構造、副作用、機能ドメインなどの特徴を抽出し、生物学的な考察を行った。クロスバリデーション実験で性能評価した結果、薬物の薬理作用プロファイルを記述子に用いることで、化学構造に依存せずに相互作用タンパク質を予測できることが確認できた。

提案手法を全ての薬物とタンパク質に対して適用し、薬物の標的タンパク質の大規模予測を行った。その予測結果の生物学的な考察や妥当性の検証を行った。薬物応答遺伝子発現プロファイルを用いる事で、薬物の作用機序の理解に繋がることを確認した。予測された薬物・タンパク質間相互作用ペアのリストを整備し、その解析結果を検索するウェブアプリケーション DINIES (Drug-target Interaction Network Inference Engine based on Supervised Analysis; <http://www.genome.jp/tools/dinies/>) を構築した。予測された標的タンパク質を通じて、薬物が本来開発された疾患とは別の疾患への効能発見の可能性を示した。

これまでの成果について、研究期間中（平成 25 年度～平成 27 年度）に、国際学術雑誌への論文発表を 10 件、学会発表を 33 件行った。その結果、国内外の多くの研究者からフィードバックを得ることができた。また、バイオインフォマティクスの学会だけでな

く、薬学、化学、医学などの学会からの招待講演や原稿執筆の依頼も増えた。本研究は創薬への応用を最終的な目的にしているが、多くの製薬会社の研究者の方にも本研究に興味を持っていただいた。現在、多くの製薬会社から共同研究のオファーが既に届いており、実際に共同研究が始まっているところもある。今後は、実験系研究者、臨床系研究者、製薬会社との共同研究を通じて新しい医薬品の開発に繋げて行きたい。

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 10 件)

- ① Hizukuri, Y., Sawada, R., and Yamanishi, Y., "Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner", *BMC Medical Genomics*, 8:82 (10 pages), 2015. 査読有り
DOI: 10.1186/s12920-015-0158-1
- ② Sawada, R., Iwata, H., Mizutani, S., and Yamanishi, Y., "Target-based drug repositioning using large-scale chemical-protein interactome data", *Journal of Chemical Information and Modeling*, 55(12), 2717-2730, 2015. 査読有り
DOI: 10.1021/acs.jcim.5b00330
- ③ Shao, Z., Hirayama, Y., Yamanishi, Y., and Saigo, H., "Mining discriminative patterns from graph data with multiple labels and its application to QSAR", *Journal of Chemical Information and Modeling*, 55(12), 2519-2527, 2015. 査読有り
DOI: 10.1021/acs.jcim.5b00376
- ④ Iwata, H., Sawada, R., Mizutani, S., Kotera, M., and Yamanishi, Y., "Large-scale prediction of beneficial drug combinations using drug efficacy and target profiles", *Journal of Chemical Information and Modeling*, 55(12), 2705-2716, 2015. 査読有り
DOI: 10.1021/acs.jcim.5b00444
- ⑤ Sawada, R., Kotera, M., and Yamanishi, Y., "Benchmarking a wide range of chemical descriptors for drug-target interaction prediction using a chemogenomic approach", *Molecular Informatics*, 33, 719-731, 2014. 査読有り
DOI: 10.1002/minf.201400066
- ⑥ Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M., and Goto, S.,

- "DINIES: drug-target interaction network inference engine based on supervised analysis", *Nucleic Acids Research*, 42, W39-W45, 2014 査読有り
DOI: 10.1093/nar/gku337
- ⑦ Yamanishi, Y., "Inferring chemogenomic features from drug-target interaction networks", *Molecular Informatics*, 32(11-12), 991-999, 2013. 査読有り
DOI: 10.1002/minf.201300079
- ⑧ Iwata, H., Mizutani, S., Tabei, Y., Kotera, M., Goto, S., and Yamanishi, Y., "Inferring protein domains associated with drug side effects based on drug-target interaction network", *BMC Systems Biology*, 7(Suppl 6):S18, 2013 査読有り
DOI: 10.1186/1752-0509-7-S6-S18
- ⑨ Tabei, Y. and Yamanishi, Y., "Scalable prediction of compound-protein interactions using minwise hashing", *BMC Systems Biology*, 7(Suppl 6):S3, 2013 査読有り
DOI: 10.1186/1752-0509-7-S6-S3
- ⑩ Kotera, M.*, Tabei, Y.*, Yamanishi, Y.*, Moriya, Y., Tokimatsu, T., Kanehisa, M., and Goto, S. (* Joint first author), "KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics", *BMC Systems Biology*, 7(Suppl 6):S2, 2013 査読有り
DOI: 10.1186/1752-0509-7-S6-S2
- [学会発表] (計 33 件)
- ① Yamanishi, Y., "Computational drug repositioning using medical big data", The 6th French-Japanese Workshop on Computational Methods in Chemistry, Kyoto, Mar.16-Mar.17, 2016.
- ② Yamanishi, Y., "Systematic drug repositioning via omics data analysis with machine learning methods", Autumn School of Chemoinformatics in Tokyo, Tokyo, Nov.25-Nov.26, 2015.
- ③ Yamanishi, Y., "Systematic drug repositioning for a wide range of diseases with machine learning methods", Study Sessions on Bioinformatics and Related Topics, Osaka, Nov.4, 2015.
- ④ Yamanishi, Y., "Systematic drug repositioning for a wide range of diseases with computational approaches", JCUP VI, Tokyo, Jun.4-Jun.5, 2015.
- ⑤ 山西芳裕 "医薬ビッグデータと機械学習によるインシリコ創薬", 新生命科学分野開拓とスーパーコンピュータ「京」, 福岡, 12/17, 2015.
- ⑥ 岩田通夫, 澤田隆介, 山西芳裕 "化合物応答遺伝子発現プロファイルの大規模解析による生理活性化合物の作用機序推定と創薬への応用", BMB2015 (第38回分子生物学会・第88回生化学会合同大会), 神戸, 12/1-4, 2015.
- ⑦ 山西芳裕 "医薬ビッグデータと機械学習によるインシリコ創薬", 日本計算機統計学会 第29回シンポジウム, 釧路, 11/27-28, 2015.
- ⑧ 山西芳裕 "医薬ビッグデータと機械学習によるインシリコ創薬", 第19回バイオメディカル研究会「人工知能のバイオメディカル応用」, 大阪, 10/30, 2015.
- ⑨ 澤田隆介, 岩田通夫, 山西芳裕 "化合物摂動と遺伝子摂動の遺伝子発現プロファイルを用いた医薬品候補化合物の標的タンパク質予測", 第4回生命医薬情報学連合大会, 宇治, 10/29-31, 2015.
- ⑩ 澤田隆介, 岩田通夫, 山西芳裕 "化合物・タンパク質間インタラクトーム情報とオミックスデータに基づくドラッグリポジショニング", 情報計算化学生物学会 2015年大会, 東京, 10/27-29, 2015.
- ⑪ 山西芳裕 "医薬ビッグデータ解析に基づくインシリコ創薬", HPCI セミナー「予測する生命科学・医療および創薬基盤」, 東京, 10/23, 2015.
- ⑫ 山西芳裕 "医療ビッグデータの創薬応用のための機械学習法", 第31回大規模データ科学に関する研究会 第1回研究集会「医療情報のコンテンツ化と統計技法の研究開発」, 札幌, 8/24, 2015.
- ⑬ Yamanishi, Y., "Analysis and inference of drug-target interaction networks", International Symposium on Bioinformatics and its Application (ISBA), Tokyo, Japan, Sep.30, 2014.
- ⑭ Yamanishi, Y., "Analysis and inference of drug-target interaction networks", The 5th French-Japanese Workshop on Computational Methods in Chemistry, Strasbourg, France, Jun.30-Jul.1, 2014.
- ⑮ Yamanishi, Y., "In silico methods for predicting drug targets and new drug indications", The 9th International Symposium of the Institute Network, Osaka, Japan, Jun.19-Jun.20, 2014.
- ⑯ Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M., and Goto, S., "DINIES: A web-based application for predicting drug-target interaction networks", The Thirteenth Asia Pacific Bioinformatics Conference

- (APBC2015), HsinChu, Taiwan, Jan. 21-Jan. 23, 2015.
- ⑰ Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M., and Goto, S., "DINIES: A web-based application for predicting drug-target interaction networks", The 22nd International Conference on Intelligent Systems for Molecular Biology (ISMB2014), Boston, USA, Jul. 11-Jul. 15, 2014.
- ⑱ 山西芳裕, "オミックスデータ解析と機械学習によるインシリコ創薬", 平成26年度九州大学高等研究院-公益財団法人九州先端科学技術研究所 研究交流会, 福岡, 12/1, 2014.
- ⑲ 山西芳裕, "薬物の潜在標的分子と新規効能のインシリコ予測", 日本化学会情報化学部会主催 第二回若手の会, 東京, 11/29, 2014.
- ⑳ 山西芳裕, "オミックスデータ解析と機械学習によるインシリコ創薬", 第23回 BMIRC 研究会, 飯塚, 10/8, 2014.
- 21 山西芳裕, "オミックスデータに基づく分子間相互作用ネットワークの予測と創薬への応用", 第3回生命医薬情報学連合大会, 仙台, 10/2-4, 2014.
- 22 山西芳裕, "ケモゲノミクスや薬理ゲノミクスに基づくインシリコ創薬", 天然薬物研究方法論アカデミー第17回シンポジウム, 富山, 7/26-27, 2014.
- 23 山西芳裕, "ネットワーク解析に基づくドラッグリポジショニング", 日本薬剤学会第29年会, さいたま, 5/20-22, 2014.
- 24 Yamanishi, Y., "Predicting drug-target interaction networks from the integration of chemical, genomic, and pharmacological spaces", International Symposium on Tumor Biology in Kanazawa & Academic Drug Discovery Symposium, Kanazawa, Japan, Jan. 23-Jan. 24, 2014.
- 25 Iwata, H., Mizutani, S., Tabei, Y., Kotera, M., Goto, S., and Yamanishi, Y., "Inferring protein domains associated with drug side effects based on drug-target interaction network", The 24th International Conference on Genome Informatics (GIW2013), Singapore, Singapore, Dec. 16-18, 2013.
- 26 Tabei, Y. and Yamanishi, Y., "Scalable prediction of compound-protein interactions using minwise hashing", The 24th International Conference on Genome Informatics (GIW2013), Singapore, Singapore, Dec. 16-18, 2013.
- 27 Kotera, M., Tabei, Y., Yamanishi, Y., Moriya, Y., Tokimatsu, T., Kanehisa, M., and Goto, S., "KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics", The 24th International Conference on Genome Informatics (GIW2013), Singapore, Singapore, Dec. 16-18, 2013.
- 28 山西芳裕, "ドラッグ・リポジショニングのためのイン・シリコ手法", 千里ライフサイエンス振興財団専門実務セミナー, 大阪, 3/20, 2014.
- 29 山西芳裕, "薬物の潜在標的分子と新規効能のインシリコ予測", 共進化社会システム創成拠点フォーラム, 東京, 3/12-13, 2014.
- 30 山西芳裕, "機械学習による薬物の標的分子や副作用の網羅的予測", 第41回構造活性相関シンポジウム, 西宮, 11/7-8, 2013.
- 31 山西芳裕, "ケモゲノミクスや薬理ゲノミクスの手法による薬物・標的タンパク質間相互作用ネットワークの予測", 第2回生命システム理論研究会, 東京, 9/13, 2013.
- 32 山西芳裕, "ゲノム情報、ケミカル情報、薬理情報から薬物標的分子を予測する統計手法", 2013年度統計関連学会連合大会, 大阪, 9/8-11, 2013.
- 33 山西芳裕, "オミックスデータから薬物標的分子を網羅的に予測するためのバイオインフォマティクス", 第3回 NGS 現場の会, 神戸, 9/4-5, 2013.

[その他]

ホームページ等

<http://www.genome.jp/tools/dinies/>

6. 研究組織

(1) 研究代表者

山西 芳裕 (YAMANISHI, Yoshihiro)

九州大学・高等研究院・准教授

研究者番号：60437267