

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 15 日現在

機関番号：84420

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25430186

研究課題名(和文) データウェアハウスとタンパク質予測情報を組み合わせた創薬標的探索システムの開発

研究課題名(英文) Development of a target discovery platform by integrating data warehousing and protein function prediction

研究代表者

水口 賢司 (Mizuguchi, Kenji)

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 バイオインフォマティクスプロジェクト・プロジェクトリーダー

研究者番号：50450896

交付決定額(研究期間全体)：(直接経費) 4,200,000円

研究成果の概要(和文)：本研究では、申請者が開発した統合データウェアハウスTargetMineと個々のタンパク質の機能や相互作用を予測する方法を統合し、創薬標的やバイオマーカー探索を効率的に実行するプラットフォームを開発することを目的とした。アミノ酸配列のみからそのタンパク質がDNAに結合するかどうかや、相互作用するタンパク質を予測する方法をヒトゲノム全体に適用し、予測によるアノテーションとTargetMineに格納している実験的なアノテーションとを統合して解析できるシステムを構築した。このシステムを呼吸器疾患の病態モデルマウスの網羅的遺伝子発現データの解析などに応用し、老化や炎症と関連する新規経路の抽出に成功した。

研究成果の概要(英文)：We have previously developed TargetMine, an integrated data warehouse for target prioritization. The current project aimed to integrate predicted annotations for protein functions and interactions into TargetMine and develop an in silico system for identifying drug targets and biomarkers. First, we have extended our methods for predicting DNA-binding proteins and DNA-binding residues, as well as protein interaction sites and interacting partners, applied these methods to the entire human proteome and constructed a system that can analyze both these predicted results and known annotations using TargetMine. We have then applied the integrated system to specific biological datasets, including mouse models for pulmonary diseases, and gained novel insights into the mechanisms of these diseases and novel pathways relevant for aging and inflammation.

研究分野：バイオインフォマティクス

キーワード：データベース 創薬ターゲット タンパク質相互作用

1. 研究開始当初の背景

多数の候補遺伝子(タンパク質)群から、次の詳細な解析(実験)に向けた少数を絞り込む問題は、現在の細胞分子生物学のあらゆる分野に共通の困難な問題と言える。そのため、遺伝子発現、配列情報、機能注釈などの情報を用い、計算機を用いて絞り込みを支援する方法が、各種提案されている[1]。研究代表者は、オープンソースの InterMine フレームワーク (<http://intermine.org>) を用いて、創薬の初期研究における支援を目的とした TargetMine データウェアハウス (<http://targetmine.mizuguchilab.org>; [2]) を開発し、C型肝炎ウイルスの増殖に関連する新規の制御因子などの予測と実験的な検証に成功してきた([3-4])。

TargetMine は、ゲノムやパスウェイ情報に加えて、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取り組んでいる点で、また独自データの組み込みや高度な意思決定を対話的に支援する枠組みを備えている点で、アカデミア、商用を問わず関連するツールとは一線を画している。しかし、標的候補遺伝子群と特に関係が深いと考えられるパスウェイや機能注釈を同定するという点で、他の手法同様、既知機能情報が少ない系では十分な結果が期待できないという限界を持つ。

一方、研究代表者を含む多数のグループにより、個々のタンパク質について、構造や機能、相互作用を予測する各種の方法が開発、提唱されている。例えば、研究代表者らはアミノ酸配列のみから立体構造を予測する FUGUE (<http://mizuguchilab.org/fugue>) や、他のタンパク質との相互作用部位を予測するソフトウェア PSIVER を開発している (<http://mizuguchilab.org/PSIVER>; [5])。これらの方法を組み合わせて、乳がん細胞中の特定のタンパク質で、がん細胞の増殖に密接に関わる部位を予測し、その結合を阻害する分子を設計することで細胞増殖を抑えることに成功した[6]。

このように、新規のバイオインフォマティクス方法論による立体構造や相互作用予測は、創薬戦略策定に貢献し得る大きな可能性を持っているが、TargetMine を中心とする創薬標的探索システムの中には、まだ組み込まれていない。これらの技術を統合することで、今までにない切り口での創薬標的の周辺分子探索が可能になると考えられる。

2. 研究の目的

本研究では、データウェアハウスとタンパク質に関する予測技術を統合し、可能な限りインシリコで創薬標的及びバイオマーカーの探索を行なう次世代創薬支援システムプロトタイプ構築を目的とした。具体的には、

相互作用予測などのデータウェアハウスへの組み込み

予測機能を取り込むためのデータウェアハウスの拡充と、既存の実験データによる基盤技術の検証

呼吸器疾患関連などの新規の実験データ解析への応用を目標とした。

3. 研究の方法

(1) タンパク質相互作用などの予測法を拡張して、TargetMine 中の既知データと予測によるアノテーションとを統合するためのシステム基盤を整備した。研究代表者と連携研究者が開発した、2つ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かどうかを予測するアルゴリズム PSOPIA や、タンパク質機能を予測する他の手法(DNA結合タンパク質やDNA結合部位の同定、酵素の詳細な機能分類など)について、ゲノムレベルでの適用が可能な形にソフトウェアを拡張し、必要なベンチマークを実行した。

(2) 既知機能情報と予測情報を組み合わせた解析を可能にするために TargetMine データウェアハウスを拡充した。TargetMine にプログラマティックにアクセスできるソフトウェアと可視化の枠組みを確立し、可能な予測アノテーションについては、予め計算した結果をデータベースに格納した。システムの検証には、公共データベースからの情報に加えて、C型肝炎ウイルス関連タンパク質発現データ、ウレタン誘導肺腫瘍モデルマウスでの遺伝子発現データなど、研究代表者らが過去に解析した現実の実験データを用いた。

(3) 共同研究者によって提供される新規の実験データを用いて、疾患関連タンパク質候補の絞り込みを行った。ここから、新規創薬標的やバイオマーカー分子の同定を目指した。

4. 研究成果

(1) タンパク質相互作用や機能予測法をゲノムレベルでの解析に拡張し、既知アノテーションとの統合基盤を構築した。

2つ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かどうかを予測する方法 PSOPIA を完成した(Murakami & Mizuguchi, *BMC Bioinformatics* 2014)。この方法を拡張して、与えられた1つのタンパク質と相互作用する可能性の高いタンパク質をデータベース中からすべて抽出する計算を可能にした。これには、TargetMine に格納されている実験的に確認されたタンパク質-タンパク質相互作用データのうち高信頼度のものを取得し、それらをトレーニングデータとして、機械学習モデルを構築した。さらに、TargetMine による既知の相互作用データと

統合的な解析ができるユーザーインターフェースを設計、実装した(図1)。

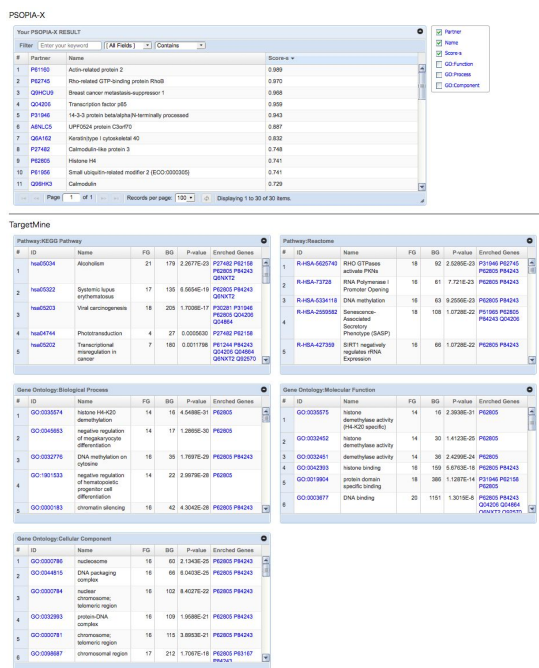


図1. データウェアハウスと統合されたタンパク質間相互作用予測システムのユーザーインターフェース

また、統合対象の一つであるタンパク質機能予測法の整備を行い、アミノ酸配列のみから酵素の詳細な機能分類(Enzyme Commission番号4桁)を同定する方法を開発した(Nagao et al., *PLoS ONE* 2014)。この方法の基礎として、タンパク質ドメインを共通な立体構造や機能を持つグループに分類する必要があり、ドメイン分類の代表的データベースであるCATHを開発している**英国 University College London のグループとの国際共同研究**を推進した(Dessailly et al., *Biochim Biophys Acta* 2013)。

さらに、タンパク質のアミノ酸配列のみから、そのタンパク質がDNA結合タンパク質であるかどうか、またどの部位でDNAと結合するかを予測する既存のアルゴリズムを拡張し、ヒトゲノム全体の予測に適用した(Ahmad et al., 論文投稿中)。この解析から、実際に新たな転写因子である可能性が高いタンパク質の同定に成功した。これら予測によるDNA結合アノテーションとTargetMineに格納している実験的なアノテーションとを統合して解析できるシステムを構築した。

(2) TargetMine データウェアハウスを拡充し、より一般的な創薬支援のためのデータ解析プラットフォームを構築した。

疾患、医薬品(開発化合物)、メカニズムの3領域で、これまでTargetMineに含まれていなかった化合物アッセイ情報や組織特異的遺伝子発現情報を取り込むとともに、遺伝子群に対して生物学的意味付けを行なう際、多数のパスウェイや機能注釈が乱立する問題を解決するため、パスウェイのクラスタリングや、結果の解釈を支援する可視化と自動化を導入した(Chen et al., *PLoS ONE* 2014)。さらに、新たな解析ワークフロー機能を実現することで、**創薬ターゲットの絞り込みを支援するシステムとして開発したTargetMineが、より一般的な創薬支援のためのデータ解析プラットフォームとして利用可能になった**(Chen et al., *Database* 2016)。定型的な解析のみでなく、多面的な意思決定を対話的に支援するTargetMineのワークフロー機能を図2に示す。

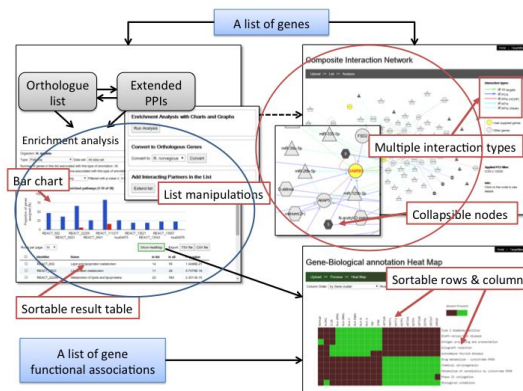


図2 TargetMineの解析ワークフロー機能

(3) 上記で開発したシステムを具体的な創薬ターゲットやバイオマーカーの探索に応用した。

上記パスウェイのクラスタリングの応用として、**米Merck社との国際共同研究**を行った。アミロイド前駆体タンパク質(Amyloid precursor protein, APP)分解の制御が、アルツハイマー病の病態理解や治療につながる期待が持たれているが、その詳細は明らかになっていない。そこで、APPの異なる分解産物をエンドポイントとしたsiRNAによる網羅的な遺伝子ノックダウンの実験結果を解析し、若年発症成人型糖尿病との関連性の具体的な分子メカニズムを提唱した(Camargo et al., *PLoS ONE* 2015)。

また、急性骨髄性白血病(AML)患者におけるWilms' tumor gene 1(WT1)特異的免疫の解析とペプチドワクチンの臨床的効果のありなしの理解に向けて、大阪大学大学院医学系研究科呼吸器・免疫アレルギー内科学癌免疫研究室との共同研究でWT1特異的細胞障害性T細胞(CTL)の単一細胞遺伝子発現解析

を行なった。WT1 の CTL effector memory 分画は遺伝子発現パターンに基づき 2 つに分けられ、ペプチドワクチンの臨床効果のありなしと関連付けられることを示した (Nakae et al., *Cancer Immunol Immunother* 2015)。

さらに、呼吸器疾患の病態モデルマウスの網羅的遺伝子発現データの解析などに応用し、老化や炎症と関連する新規経路の抽出に成功した (論文準備中)。

<引用文献>

- [1] Moreau, Y. and L.C. Tranchevent, (2012) *Nat Rev Genet*, 13, 523-36.
- [2] Chen, Y.A., L.P. Tripathi, and K. Mizuguchi, (2011) *PLoS One*, 6, e17844.
- [3] Tripathi, L.P., et al., (2010) *Mol Biosyst*, 6, 2539-53.
- [4] Tripathi, L.P., et al., (2012) *J Proteome Res*, 11, 3664-79.
- [5] Murakami, Y. and K. Mizuguchi, (2010) *Bioinformatics*, 26, 1841-8.
- [6] Yoshimaru et al., (2013) *Nature commun*, 4:2443.

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 2 2 件)

Chen Y. A., Tripathi L. P., Mizuguchi K.、An integrative data analysis platform for gene set analysis and knowledge discovery in a data warehouse framework、Database (Oxford)、査読有、2016、DOI: 10.1093/database/baw009

Tsuchiya Y., Mizuguchi K.、The diversity of H3 loops determines the antigen-binding tendencies of antibody CDR loops、Protein Sci、査読有、25(4)、2016、815-25、DOI: 10.1002/pro.2874

Prathipati P., Mizuguchi K.、Integration of Ligand and Structure Based Approaches for CSAR-2014、J. Chem. Inf. Model、査読有、2015、Nov 5. [Epub ahead of print]、DOI: 10.1021/acs.jcim.5b00477

Nagao C., Nagano N., Mizuguchi K.、Prediction of detailed enzyme functions and identification of specificity determining residues by random forests、PloS One、査読有、9(1)、2014、e84623、DOI: 10.1371/journal.pone.0084623

Murakami Y.、Mizuguchi K.、Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators、BMC

bioinformatics、査読有、15(1)、2014、213、DOI: 10.1186/1471-2105-15-213

Chen Y. A., Murakami Y.、Ahmad S., Yoshimaru T., Katagri T., Mizuguchi K.、Brefeldin a-inhibited guanine nucleotide-exchange protein 3 (BIG3) is predicted to interact with its partner through an ARM-type alpha-helical structure、BMC Res Notes、査読有、7(1)、2014、435、DOI: 10.1186/1756-0500-7-435

Chen Y. A., Tripathi L. P., Dessailly B. H., Nyström-Persson J., Ahmad S., Mizuguchi K.、Integrated Pathway Clusters with Coherent Biological Themes for Target Prioritisation Blockade of TLR3 protects mice from lethal radiation-induced gastrointestinal syndrome、PLoS One、査読有、9(6)、2014、e99030、DOI: 10.1371/journal.pone.0099030

Dessailly B. H., Dawson N. L., Mizuguchi K.、Orengo C. A.、Functional Site Plasticity in Domain Superfamilies、BBA - Proteins and Proteomics、査読有、1834(5)、2013、874-89、DOI: 10.1016/j.bbapap.2013.02.042

Tripathi L. P., Kambara H., Chen Y. A., Nishimura Y., Moriishi K., Okamoto T., Morita E., Abe T., Mori Y., Matsuura Y., Mizuguchi K.、Understanding the Biological Context of NS5A-Host Interactions in HCV Infection: A Network-Based Approach、Journal of Proteome Research、査読有、12(6)、2013、2537-51、DOI: 10.1021/pr3011217

Nystrom J., Igarashi Y., Ito M., Morita M., Nakatsu N., Yamada H., Mizuguchi K.、Toxygates: interactive toxicity analysis on a hybrid microarray and linked data platform、Bioinformatics、査読有、29(23)、2013、3080-6、DOI: 10.1093/bioinformatics/btt531

[学会発表](計 4 4 件)

水口賢司、構造からネットワークへ：計算生物学による創薬研究、日本学術振興会構造生物 169 委員会 第 4 9 回研究会、東京、2016.3.9 (招待講演)

水口賢司、データ統合と計算の融合による創薬研究、バイオスーパーコンピューティング神戸 2015、神戸、2015.11.30 (招待講演)

水口賢司、データウェアハウスによる創薬関連データ統合と解析の実際、第 361 回 CBI 学会研究講演会、大阪、2015.4.24 (招待

講演)

水口賢司、乳癌細胞増殖の分子機構解明から創薬へ：タンパク質構造と相互作用予測による主導、CBI 学会 2014 年大会，東京，2014.10.29 (招待講演)

水口賢司、Data integration and protein network analysis for early stage drug discovery、Structural Life Science 7th International Conference on Structural Genomics (ICSG2013-SLS)，札幌，2013.7.31 (招待講演)

〔図書〕(計1件)

水口賢司、南山堂、計算生物学によるシステムの理解から創薬へ、生命科学から創薬へのイノベーション、1巻、2014、82-88

〔その他〕

TargetMine トップページ

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

TargetMine 解析ワークフロー機能

<http://targetmine.mizuguchi lab.org/auxiliary-toolkit>

TargetMine パスウェイクラスタリング機能

<http://targetmine.mizuguchi lab.org/integrated-pathway-clusters>

PSOPLA- アミノ酸配列のみからの相互作用タンパク質の予測プログラム

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

研究室の研究紹介ページ

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

6. 研究組織

(1) 研究代表者 水口 賢司 (MIZUGUCHI, Kenji) 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・プロジェクトリーダー 研究者番号：50450896

(2) 研究分担者

なし

(3) 連携研究者 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・プロジェクト研究員 研究者番号：20548424