

令和 2 年 7 月 7 日現在

機関番号：84420

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K07268

研究課題名(和文) データウェアハウスとシグナル伝達シミュレーションによる創薬標的予測能力の向上

研究課題名(英文) Improved identification of drug targets by data warehousing and simulations of signal transduction pathways

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 AI健康・医薬研究センター・センター長

研究者番号：50450896

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：創薬ターゲットの探索や病態解析において、遺伝子、タンパク質、代謝物などを網羅的に測定したオミックスデータや診療情報、医薬品情報などを統合して多面的に解析する必要性が増している。報告者らが開発した創薬支援ツールであるデータウェアハウスTargetMineに、本研究によって、従来の遺伝子、タンパク質解析に加えて、代謝物の網羅的測定データを用いて、それらの相互関係、制御関係を導き出せる機能を付与した。またタンパク質間相互作用の確からしさを分類してスコア化することにより、より現実的なパスウェイを推測することが可能となった。

研究成果の学術的意義や社会的意義

創薬研究において、開発が容易な創薬ターゲットに作用する薬剤はすでに研究開発し尽くされ、現在残っているのは高難易度の創薬ターゲットのみであること、また難病などにおいて、まだ真の標的が見いだされていないことなどから、「創薬ターゲットの枯渇」が深刻化している。この課題に対応するため、ヒトゲノム等の生体分子のデータと臨床データや医学、生物学的データ、医薬品関連のデータなどとの統合的な解析により、疾患病態の解明や真の創薬ターゲットの発見につなげることが求められている。本研究成果は、この統合的な解析のツールを提供し、真の創薬ターゲットを導き出すための支援につながるものである。

研究成果の概要(英文)：In drug discovery and disease biology, the rapid proliferation of high-throughput 'omics' data has necessitated a need for tools that allow the researchers to combine and analyze different types of biological data. TargetMine is an integrative data analysis platform for target prioritization and broad-based biological knowledge discovery. In this study, we added new functions to TargetMine, treating new biological data types and enhanced visual and analytical features. First, to facilitate a more effective metabolomics analysis with TargetMine, we integrated the compound database KEGG COMPOUND, as well as KEGG reactions. As a result, given a list of compounds, the users can retrieve the enzymatic reactions involving these metabolites and map them to the corresponding pathways. Second, we evaluated all protein-protein interactions (PPIs) in TargetMine using our own in silico prediction method, and individual PPIs were tagged with scores of new attributes of reliability.

研究分野：バイオインフォマティクス

キーワード：トランスオミックス解析 代謝物情報 統合データベース KEGGパスウェイ エンリッチメント解析
タンパク質間相互作用

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

創薬研究において、創薬ターゲット（医薬品が作用するタンパク質などの生体分子）が枯渇している状態であると言われて久しい。それは、開発が容易な創薬ターゲットに作用する薬剤はすでに研究開発し尽くされ、現在残っているのは高難易度の創薬ターゲットのみであること、また難病などにおいて、まだ真の標的が見いだされていないことが理由であると考えられている。この深刻化する「創薬ターゲットの枯渇」に対応するためには、ヒトゲノム等の生体分子のデータと臨床データや医学、生物学的データ、医薬品関連のデータなどとの統合的な解析手法を取り入れる必要が生じてきた。すなわち、多種多様な医療情報データが指数的に増大している中、これらのデータを統合し、因果関係を緻密に解析することにより、疾患病態の解明や真の創薬ターゲットの発見につなげることが求められてきた。さらに、次世代シーケンサーや質量分析器などの測定技術の進歩により、遺伝子、タンパク質、代謝物を網羅的に測定し、それらを解析するオミックス解析が普及し、ビッグデータが蓄積し始めた。これらのデータをもとに生体内で起こる、あるいは疾患状態において起こる生物学的変化を捉えるためには、上記の多種多様な情報の統合に加えて、これらの網羅的解析データを多面的に解析するマルチオミックス解析あるいは、多層的に相互の関連性をつなげて解析するトランスオミックス解析の必要性が増してきた。

報告者らは、遺伝子関連データベースである NCBI Gene (<https://www.ncbi.nlm.nih.gov/gene>) やタンパク質関連データベースの Uniprot (<https://www.uniprot.org>) など国際的に広く使用されている主要な 30 個以上の公共データベースからのデータを統合したデータウェアハウス TargetMine (<https://targetmine.mizuguchilab.org>) を開発してきた。TargetMine は、タンパク質立体構造や医薬品関係データ、転写因子とその作用遺伝子の関係等のデータを深くかつ正確に関連付けて検索することが可能なシステムであり、本助成事業の 2013-2015 においては、相互作用するタンパク質を予測するシステムとの統合を行い、新規の実験によるデータ取得への依存度を減らし、可能な限りインシリコで創薬ターゲットの探索を行うプラットフォームのプロトタイプを構築した。このシステムを利用することで、複数の候補遺伝子または候補タンパク質から、特定の機能や疾患との関連性を有する遺伝子又はタンパク質を絞り込むことが可能となった。

しかし、上述のマルチオミックス解析あるいはトランスオミックス解析を可能にするためには、代謝物情報のマイニング機能を導入し、遺伝子、タンパク質および代謝物の網羅的解析データからそれらの相互関係、制御関係を導き出せる機能を付与する必要がある。さらに、生体内の反応をより正確に把握するためには、タンパク質相互作用予測をもとにしたより現実的なシグナルネットワークを再構築し、創薬ターゲットの探索・同定手法を精緻化する必要がある。

2. 研究の目的

本研究は、データウェアハウス TargetMine に網羅的代謝物測定データのマイニング機能を付与するとともに、可視的に、また効率的にマルチオミックス、トランスオミックス解析が行えるシステムを構築すること、タンパク質相互作用予測機能を用いてより現実的なシグナルネットワークを再構築すること、予測能を向上させることにより、シミュレーション技術へつなげることを目的とした。これらの機能を新たに付与することにより、多種多様な情報の統合に加えて、遺伝子、タンパク質、代謝物の網羅的解析データを多面的、多層的に解析するマルチオミックス解析、トランスオミックス解析へも対応可能となり、新規創薬ターゲットの探索や疾患特異的なバイオマーカー探索へのより有用な手法を提供できると考えた。

3. 研究の方法

TargetMine への代謝物データのマイニング機能およびパスウェイへのマッピング機能とのリンクを付与するために、KEGG などライフサイエンスのデータベースに含まれる化合物情報を取り込み、他のデータベースと統合し、さらにプログラミングで操作しつつ、そのデータの扱いを補助するスクリプトを作成した。かつ代謝物データの代謝パスウェイ上へのマッピング、エンリッチメントをはじめとする各種統計解析手法も実装した。本システムの検証は、遺伝子、タンパク質、代謝物の網羅的解析データを既報より取得して実施した。さらに、パスウェイを反応経路ごとに分割して、タンパク質間相互作用のエビデンスをもとに、既存の相互作用の有無の明示化、および現実的なネットワークの再構築を行った。

4. 研究成果

本研究では、まず第一に TargetMine で、代謝物・化合物の情報を統合することにより、すでに統合していた遺伝子、タンパク質情報と同様に、たとえばエンリッチメント解析を行うことができるようになった。さらに TargetMine をプログラミングで操作しつつ、そのデータの扱いを補助するスクリプトを作成することで、より大量のデータを高速で処理できるようになり、各

種のおミックスデータの可視化・簡便化を実現した。



図1 TargetMineを用いる生物学的データベース検索の簡便化

具体的には、メタボローム解析・プロテオーム解析・トランスクリプトーム解析データに記述される分子のIDとコントロールに対する変化度合いの数値のデータをリストとし、TargetMineへの検索入力とする。以降はワークフローに従って、エンリッチメント解析の結果のテキストファイル保存と、上位のパスウェイに対して入力した分子のIDを強調表示したマッピング（例えば特定のパスウェイについて、KEGGパスウェイへのマッピングや、グローバルマップへのマッピング）が行える。

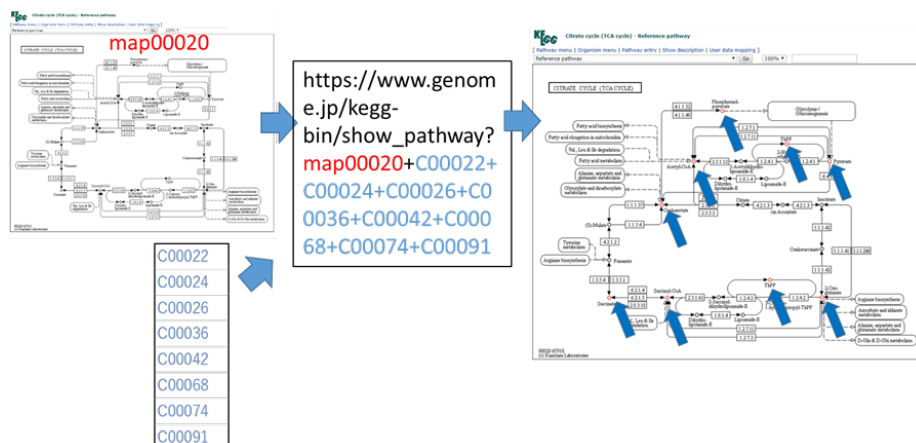


図2 KEGGパスウェイに発現変動分子をマッピング

また化合物を制御する酵素の情報を引き出し、酵素をコードする遺伝子の情報を引き出すワークフローを作成することで、代謝物データをマイニングしたのちに、酵素をコードする遺伝子の発現情報や関連疾患の情報がTargetMineにリンクされていることから、代謝物データを出発点とした創薬ターゲット探索も行えるようになった。

これらは、TargetMineの対話的な検索システムであるクエリービルダーに組み込み、横断検索を可能とした。

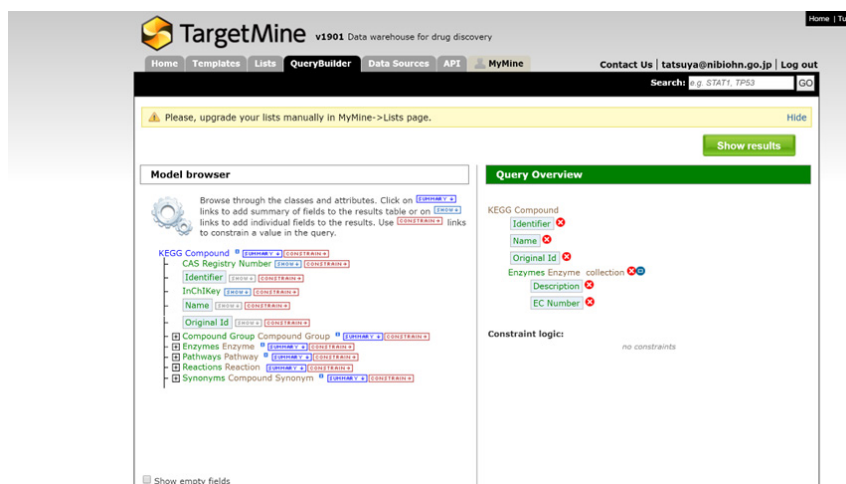


図2 TargetMineにおけるクエリービルダー例

既報の遺伝子、タンパク質、代謝物のそれぞれのオミックスデータ(Williams, E. G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C. A., et al. Systems proteomics of liver mitochondria function. Science 352, aad0189(2016))を用いて、本システムの検証を行った。すなわち、各オミックスデータを用いてエンリッチメント解析等を行った結果、遺伝子、タンパク質、代謝物のそれぞれのオミックスから共通のパスウェイがエンリッチされ、それらのパスウェイを可視的に確認するとともに、さらに脂質代謝に関連する重要な代謝パスウェイを新たに示唆することができた(Chen Y-A, et al., Front. Genet. 10:934. 2019)。以上のことから、遺伝子、タンパク質に加えて、代謝物データのマイニングから解析までを行える機能を付与することで、マルチオミックス解析、トランスオミックス解析への手法を提供することが可能となった。

次に、本研究において、タンパク質間相互作用の信頼性を実験的に得られたデータに裏付けられているか否かについて、TargetMineに統合したすべてのタンパク質間相互作用データを調査し、あたらしい信頼性の属性を付与した。具体的には、報告者らが、以前、開発したPSOPIAシステム、すなわち、実験的に裏付けられたタンパク質相互作用に基づいて、タンパク質のペアの間の相互作用の可能性を予測するシステムを利用して、出力されたPSOPIAスコアを新しい属性PsopiaScoreとして、個々のタンパク質相互作用データにタグ付けした。本実装により、ユーザーは、遺伝子とタンパク質が相互作用するパートナーや関心のある遺伝子やタンパク質のリストにクエリーを実行して、十分に信頼性が高いと判断したすべての相互作用データで構成されるネットワークを推測することができるようになった。

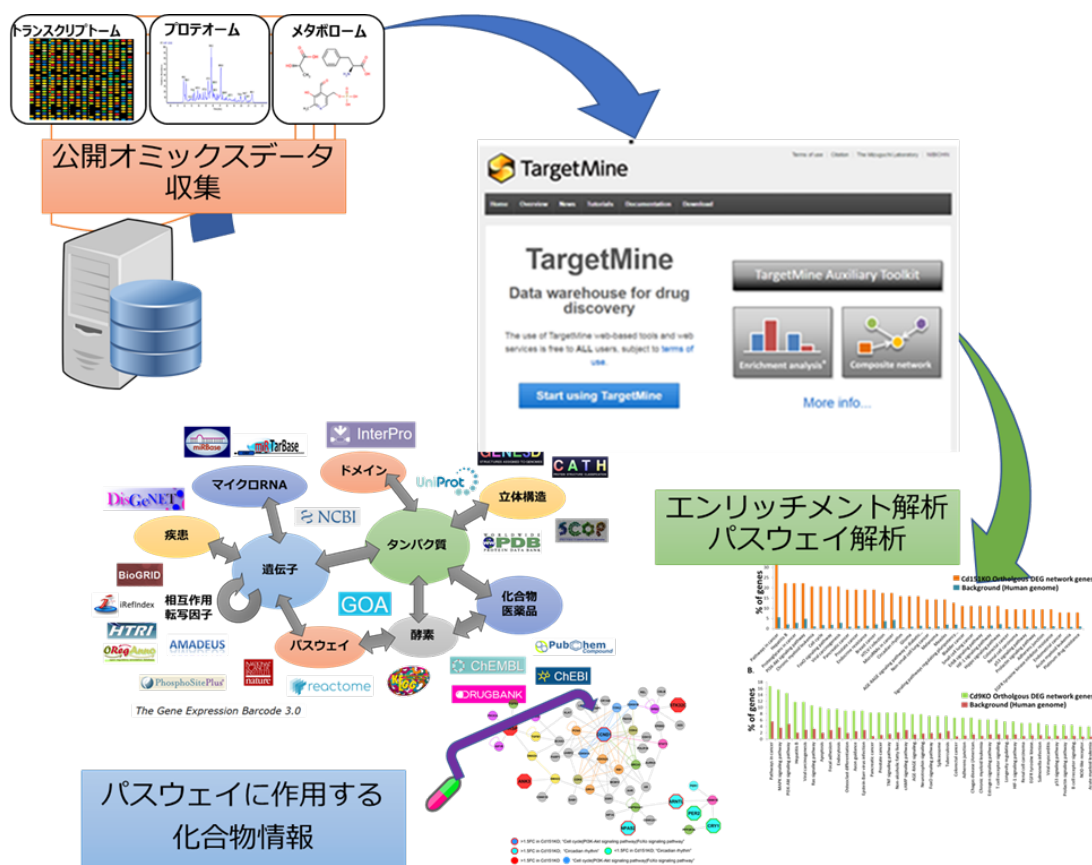


図4 本研究において実現したマルチオミックス、トランスオミックス解析が可能となったTargetMine

5. 主な発表論文等

〔雑誌論文〕 計13件（うち査読付論文 8件 / うち国際共著 1件 / うちオープンアクセス 9件）

1. 著者名 Tripathi L.P., Esaki T., Itoh M., Chen Y.A, Mizuguchi K.	4. 巻 3
2. 論文標題 Integrative Analysis of Multi-Omics Data	5. 発行年 2019年
3. 雑誌名 Encyclopedia of Bioinformatics and Computational Biology	6. 最初と最後の頁 194-199
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/B978-0-12-809633-8.20096-4	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Chen Y.A., Yogo E., Kurihara N., Ohno T., Higuchi C., Rokushima M., Mizuguchi K.	4. 巻 8
2. 論文標題 Assessing drug target suitability using Targetmine	5. 発行年 2019年
3. 雑誌名 F1000Research	6. 最初と最後の頁 233
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 夏目やよい, 相崎健一, 北嶋聡, 陳怡安, 水口賢司, 菅野純	4. 巻 43
2. 論文標題 TargetMineによる標的予測	5. 発行年 2018年
3. 雑誌名 The Journal of Toxicological Sciences	6. 最初と最後の頁 S76
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 長尾知生子, 夏目やよい, 水口賢司	4. 巻 1(1)
2. 論文標題 創薬における計算機の果たす役割 -プレジジョンメディシンに向けて-	5. 発行年 2018年
3. 雑誌名 Precision Medicine	6. 最初と最後の頁 28-31
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 伊藤眞里	4. 巻 54(9)
2. 論文標題 バイオメディカル・基礎から臨床への開発プロセス(1)~人工知能を用いる創薬テーマ創出	5. 発行年 2018年
3. 雑誌名 医薬ジャーナル	6. 最初と最後の頁 2029-2032
掲載論文のDOI(デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Chen Y.A., Tripathi L.P., Mizuguchi K.	4. 巻 印刷中
2. 論文標題 Data Warehousing with TargetMine for Omics Data Analysis	5. 発行年 2019年
3. 雑誌名 Methods Molecular Biology 1986(Microarray Bioinformatics)	6. 最初と最後の頁 印刷中
掲載論文のDOI(デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 Murakami Yoichi, Mizuguchi Kenji	4. 巻 -
2. 論文標題 PSOPIA: Toward more reliable protein-protein interaction prediction from sequence information	5. 発行年 2017年
3. 雑誌名 IEEE Conferences	6. 最初と最後の頁 255 ~ 261
掲載論文のDOI(デジタルオブジェクト識別子) 10.1109/ICIBMS.2017.8279749	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 Murakami Yoichi, Mizuguchi Kenji	4. 巻 3
2. 論文標題 Making protein-protein interaction prediction more reliable with a large-scale dataset at the proteome level	5. 発行年 2017年
3. 雑誌名 Journal of Bioinformatics and Neuroscience	6. 最初と最後の頁 91 ~ 98
掲載論文のDOI(デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ahmad Shandar, Prathipati Philip, Tripathi Lokesh P, Chen Yi-An, Arya Ajay, Murakami Yoichi, Mizuguchi Kenji	4. 巻 46
2. 論文標題 Integrating sequence and gene expression information predicts genome-wide DNA-binding proteins and suggests a cooperative mechanism	5. 発行年 2018年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 54~70
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/nar/gkx1166	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Jin Yingji, Takeda Yoshito, Kondo Yasushi, Tripathi Lokesh P., Tetsumoto Satoshi, Tsujino Kazuyuki, Kuroyama Muneyoshi, Iwahori Kota, Hirata Haruhiko, Takimoto Takayuki, Suzuki Mayumi, Nagatomo Izumi, Sugimoto Ken, Fujii Yuta, Kida Hiroshi, Mizuguchi Kenji, Ito Mari	4. 巻 8
2. 論文標題 Double deletion of tetraspanins CD9 and CD81 in mice leads to a syndrome resembling accelerated aging	5. 発行年 2018年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 5145
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-018-23338-x	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tripathi Lokesh P., Murakami Yoichi, Chen Yi-An., Mizuguchi Kenji	4. 巻 3
2. 論文標題 Network-based analysis for biological discovery	5. 発行年 2018年
3. 雑誌名 Encyclopedia of Bioinformatics and Computational Biology	6. 最初と最後の頁 283~291
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/B978-0-12-809633-8.20674-2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 伊藤 眞里、長尾 知生子、夏目 やよい、水口 賢司	4. 巻 35
2. 論文標題 公共データベースの統合からAIによるデータマイニングに向けて	5. 発行年 2017年
3. 雑誌名 日本化学会情報化学部会誌	6. 最初と最後の頁 210~213
掲載論文のDOI (デジタルオブジェクト識別子) 10.11546/cicsj.35.210	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 水口 賢司	4. 巻 22
2. 論文標題 創薬の初期研究におけるデータベース構築とモデリング	5. 発行年 2017年
3. 雑誌名 学術の動向	6. 最初と最後の頁 7_62~7_65
掲載論文のDOI (デジタルオブジェクト識別子) 10.5363/tits.22.7_62	査読の有無 無
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計20件(うち招待講演 17件/うち国際学会 3件)

1. 発表者名 Kameyama T., Chen Y.A., Nojima Y., Fujiwara T., Esaki T., Tripathi L., Ito M., Mizuguchi K.
2. 発表標題 Adding trans-omics analysis features to TargetMine
3. 学会等名 C B I学会2018年 大会
4. 発表年 2018年

1. 発表者名 水口賢司, 坂手龍一, 五十嵐芳暢, 陳怡安, 樋口千洋, 長尾知生子, 深川明子
2. 発表標題 創薬・健康・栄養研究を支援する支援するNIBIOHNのデータベース
3. 学会等名 第41回日本分子生物学会年会(特別企画バイオデータベースコーナー)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 タンパク質インフォマティクスのためのデータ統合
3. 学会等名 第18回日本蛋白質科学会(招待講演)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 創薬におけるAI活用
3. 学会等名 C B I 学会 (情報計算化学生物学会) 2018年大会 (プレナリー講演) (招待講演)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 A I 開発基盤としてのデータ統合：薬物動態モデリングとターゲット探索
3. 学会等名 第 1 回日本メディカルA I 学会学術集会 (招待講演)
4. 発表年 2019年

1. 発表者名 水口賢司
2. 発表標題 Data integration and computational systems approaches towards rational drug discovery
3. 学会等名 InCoB 2018 (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Mizuguchi K.
2. 発表標題 Data integration as a basis for computational modelling for rational drug discovery
3. 学会等名 The 18th KIAS Conference on Protein Structure and Function (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 Data integration and computational systems approaches to target discovery and pharmacokinetic modelling
3. 学会等名 日英セミナー Big Data and AI in Drug Discovery: Resources and Developments (招待講演)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 Data integration and computational modelling for drug discovery
3. 学会等名 第6回ITbM国際シンポジウム(招待講演)(国際学会)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 創薬・疾患研究のためのデータ統合の実際
3. 学会等名 トーゴの日シンポジウム2018(招待講演)
4. 発表年 2018年

1. 発表者名 伊藤真里
2. 発表標題 AIと創薬
3. 学会等名 NPO法人J-Win Next Stage 『第四次産業革命研究会』(招待講演)
4. 発表年 2018年

1. 発表者名 水口賢司
2. 発表標題 データ統合と人工知能技術による新薬創出の加速
3. 学会等名 彩都産学官連携フォーラム 2019 (招待講演)
4. 発表年 2019年

1. 発表者名 水口賢司
2. 発表標題 計算生物学と創薬のためのデータ統合
3. 学会等名 シンポジウム「理論生物物理学の現在と未来」(招待講演)
4. 発表年 2019年

1. 発表者名 水口賢司
2. 発表標題 インフォマティクスの活用による新薬創出の加速を目指して
3. 学会等名 第5回インターフェックス大阪(招待講演)
4. 発表年 2019年

1. 発表者名 水口賢司
2. 発表標題 データ統合と計算システム生物学による創薬
3. 学会等名 バイオインタラクション研究会 第4回ワークショップ(招待講演)
4. 発表年 2019年

1. 発表者名 水口賢司
2. 発表標題 タンパク質立体構造の比較とデータベース
3. 学会等名 第386回CBI学会講演会（招待講演）
4. 発表年 2017年

1. 発表者名 水口賢司
2. 発表標題 多階層データの統合モデリングによる薬物動態予測 システムの構築
3. 学会等名 CBI学会2017年大会フォーカストセッション（招待講演）
4. 発表年 2017年

1. 発表者名 水口賢司
2. 発表標題 Data integration and statistical/ab initio modeling towards rational drug discovery
3. 学会等名 The 55th Annual Meeting of the Biophysical Society of Japan（招待講演）
4. 発表年 2017年

1. 発表者名 水口賢司, 坂手龍一, 深川明子, 五十嵐芳暢, 陳怡安, 樋口千洋, 長尾知生子
2. 発表標題 創薬・健康・栄養 研究を支援するNIBIOHNのデータベース
3. 学会等名 2017年度生命科学系学会合同年次大会（ConBio2017）
4. 発表年 2017年

1. 発表者名 水口賢司
2. 発表標題 データ統合と多階層モデリングによる創薬の効率化
3. 学会等名 徳島大学研究クラスター 統合的がん創薬研究クラスター 特別セミナー（招待講演）
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	伊藤 眞里 (Ito Mari) (50728102)	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 AI健康・医薬研究センター・上席研究員 (84420)	