

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 13 日現在

機関番号：32612

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K08689

研究課題名(和文)メタボロミクスとネットワーク生物学を用いた新規腎臓病関連遺伝子予測AIの開発

研究課題名(英文)Development of AI for the discovery of novel kidney disease genes using metabolomics and network biology

研究代表者

齋藤 輪太郎 (SAITO, Rintaro)

慶應義塾大学・政策・メディア研究科(藤沢)・特任教授

研究者番号：40348842

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：メタボローム分析装置から得られる大量の解釈困難部分(未知ピーク)の有効利用の可能性を探り、また患者の代謝物質のパターンから急性腎症(AKI)の発症予測を行った。前者では、鶴岡メタボロームコホート研究参加者約3,000人分のメタボロームデータを活用し、生体内の代謝物質を反映すると思われる未知ピークを抽出するためのパイプラインの構築を行った。後者では、AKIを発症した患者の尿に特有の代謝物質の変化を発見できないか調べるべく、手術後に集中治療室に入室した患者121人の、未知ピークを含む尿中メタボロームの分析を行い、さらに機械学習を用いて、その後患者がAKIを発症するかを比較的高い精度で予測した。

研究成果の学術的意義や社会的意義

代謝物質は生命活動の源であると同時に、生命活動の結果合成されるものでもあり、細胞内でどのような現象が起きているかを探る大きな手がかりとなる。細胞内の代謝物質の網羅的分析、すなわちメタボローム分析の応用分野は、分子生物学のような基礎研究にとどまらず、医学や食品学を含めた健康科学にまで広がっており、ガンなどの疾患進行のメカニズムの解明や、新規バイオマーカーの発見等、その成果は多岐に渡る。本研究成果は、メタボロームデータを効果的に情報処理し、検体に関するより多くの情報抽出を可能にするものである。機械学習を活用し、急性腎症発症の予測を比較的高い精度で行えたことでその医学的有用性が示唆された。

研究成果の概要(英文)：We explored the possibility of effective utilization of a large amount of uncharacterized part of metabolome data (uncharacterized peaks) obtained from analytical instruments, and predicted the onset of acute kidney injury (AKI) based on metabolome profiles of the patients. In the former, we used metabolome data from ~3,000 participants in the Tsuruoka Metabolome Cohort Study to construct a pipeline for extracting uncharacterized peaks that are likely to reflect actual metabolites in the body. In the latter, we analyzed the urinary metabolites including uncharacterized peaks in the urine samples of 121 patients who entered the intensive care unit after surgery, to investigate whether changes in metabolome profiles specific to patients who developed AKI could be detected. Machine learning enabled us to predict onset of AKI with reasonably high accuracy.

研究分野：メタボロミクス

キーワード：メタボロミクス 腎臓病 機械学習

1. 研究開始当初の背景

腎臓病患者は我が国に約 1300 万人いると見積もられており、精度の高い腎臓病診断・罹患リスク予測のためのバイオマーカーの発見や、疾患の分子メカニズムの全容解明等が喫緊の課題となっている。そのための方法論として、生体内の代謝物質を網羅的に分析する「メタボロミクス」が注目を集めている。

2. 研究の目的

1 検体当たり数百メガバイト(MB)以上に及ぶメタボローム生データのなるべく多くの部分を有効活用するためのデータ処理方法を開発し、その信頼性を評価すること、ならびに、機械学習等を用いて、メタボロームデータから腎臓病罹患リスクを予測する方法を開発することを研究目的とした。

3. 研究の方法

本研究では、腎臓病患者の血液や尿サンプルより得られたメタボロームデータを、機械学習等の AI (Artificial Intelligence, 人工知能) 技術を活用しつつ、効果的に情報解析し、バイオマーカー候補を探索したり、疾患進行に関わる分子メカニズムの考察を支援する手法の開発を行った。これは、メタボロームの生データの効果的な処理方法の開発と、実際の腎臓病患者の尿メタボローム解析の 2 つで構成され、特に後者は研究協力機関である名古屋大学大学院医学系研究科と取り組んだ。

前者では、メタボローム分析装置(CE-TOFMS)から得られる大量の、通常は注目の対象外とされる、「未知ピーク」の有効利用の可能性を探った(Saito R *et al. J Clin Med* 2021)。その方法として、未知ピークを自動抽出し、様々な指標でその信頼性を評価するパイプラインの構築を行った(図 1b)。

そして上記結果を踏まえた上で、次に急性腎障害(AKI)患者に特有の代謝物質の変化を発見すべく、手術後に集中治療室(ICU)に入室した 121 名の患者(そのうち、58 名はその後 AKI を発症)の「未知ピーク」を含む尿中メタボロームの時系列データから AKI 発症予測を行った(Saito R *et al. Metabolites* 2021)。予測の方法として、LASSO と呼ばれる機械学習を用いた。

4. 研究成果

(1) 血液・尿のメタボローム生データ中に含まれる未知ピークの信頼性解析

本研究では、メタボローム分析装置から得られる大量の未知ピークの有効利用の可能性を探った。図 1a のように、分析装置からは、大量のピークと呼ばれる波形が得られるが、それらがどの代謝物質に対応するかが特定できるのは一部である。残りのピーク、すなわち「未知ピーク」(図 1a で“?”マーク付のピーク)はただのノイズとの区別や生化学的解釈が難しいこともあり、大抵は深く調べられることなく捨てられてしまうのが現状である。しかしながら過去には、血液等の生体サンプルの由来個人の病態と高い相関を示す未知ピークも当研究所により発見されており(Hirayama A *et al. Anal Bioanal Chem* 2012 等)、様々な検体のメタボロームデータを網羅的に調べれば、健康状態や疾患発症リスクを反映するような貴重な未知ピークが他にも多数埋もれている可能性が十分にある。

そこで、既に測定済みの鶴岡メタボロームコホート研究(TMCS)のベースラインのメタボロームデータ(約 10,000 人分)の 1/3 を活用し、未知ピークの信頼性を評価するパイプラインの構築を行った。より具体的には、パイプラインの中で、コホートデータに含まれる各ピークの検出率、ピークのシグナル強度、サンプルを希釈した時のピーク形状の変化、ピーク同士の相関等、様々な評価項目を組み合わせ、これを用いて未知ピークと既知ピークに対する網羅的評価を行った(図 1b)。

また、未知ピークを処理する上で欠かせなかったのは、ピーク処理の自動制御である。大規模コホートに含まれる大量の未知ピークを処理するには、人手ではとても追いつかないため、この部分の自動化が必須であった。そこで、当研究所で蓄積した自動化のノウハウをパイプラインに投入し、併せて評価を行った。

その結果、血液中に含まれるピークの74%、尿に含まれるピークの95%について、高い精度でピークを自動抽出していることが観測された(手動でピークを抽出した場合と比較して、ピーク面積の相関係数が0.8以上)。また、今回の条件では、自動検出された血中未知ピーク276個、尿中未知ピーク202個のうち、241個(血中)・128個(尿中)はノイズやピークの重複の可能性が高く、残りが生体内代謝物質に対応している可能性が示唆された。図1cは生体内代謝物質に対応していると推定された血中未知ピークの一例である。

ここに、未知ピークを有効利用するための基盤が整えられたが、本研究には、パイプラインの改良・評価項目の見直しや、様々な条件検討、性能の最適化、ピーク自動処理精度の改良等、数多くの課題が残っている。しかしながら、検体の未知ピークから今まで埋もれていた健康情報を引き出し、今まで難しかった疾患の検査やリスク判定を行う技術確立へ向けた大きな一歩となった。

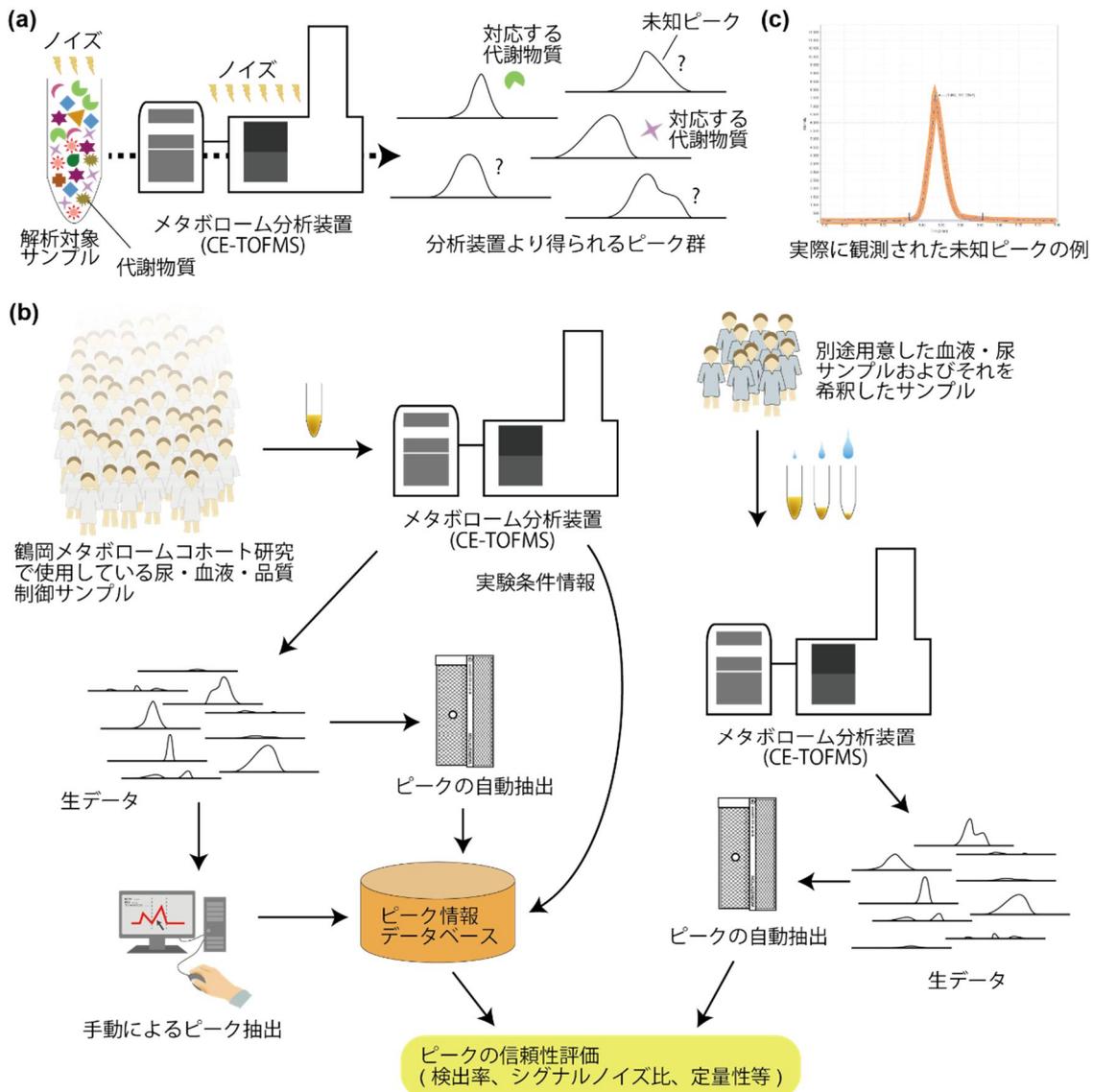


図1: 未知ピーク抽出のためのパイプラインの概要。(a) サンプルの分析より得られるピーク群と未知ピーク (b) 未知ピーク抽出・評価のためのパイプライン。(c) 実際に観測された未知ピークの例。

(2) 急性腎障患者の尿中メタボロームの時系列解析

急性腎障害 (Acute kidney injury, AKI) は、数時間以内に起こる腎機能の急速な低下であり、その原因は侵襲性の高い手術の後の急速な血流の低下や、鎮痛剤の過剰摂取等様々である。AKIは時に命に関わるため、発症が疑われる場合には、速やかに緊急の治療体制を整える必要があるが、その正確な早期予測は、現在知られているタンパク質のバイオマーカー候補等を用いても依然として困難である。

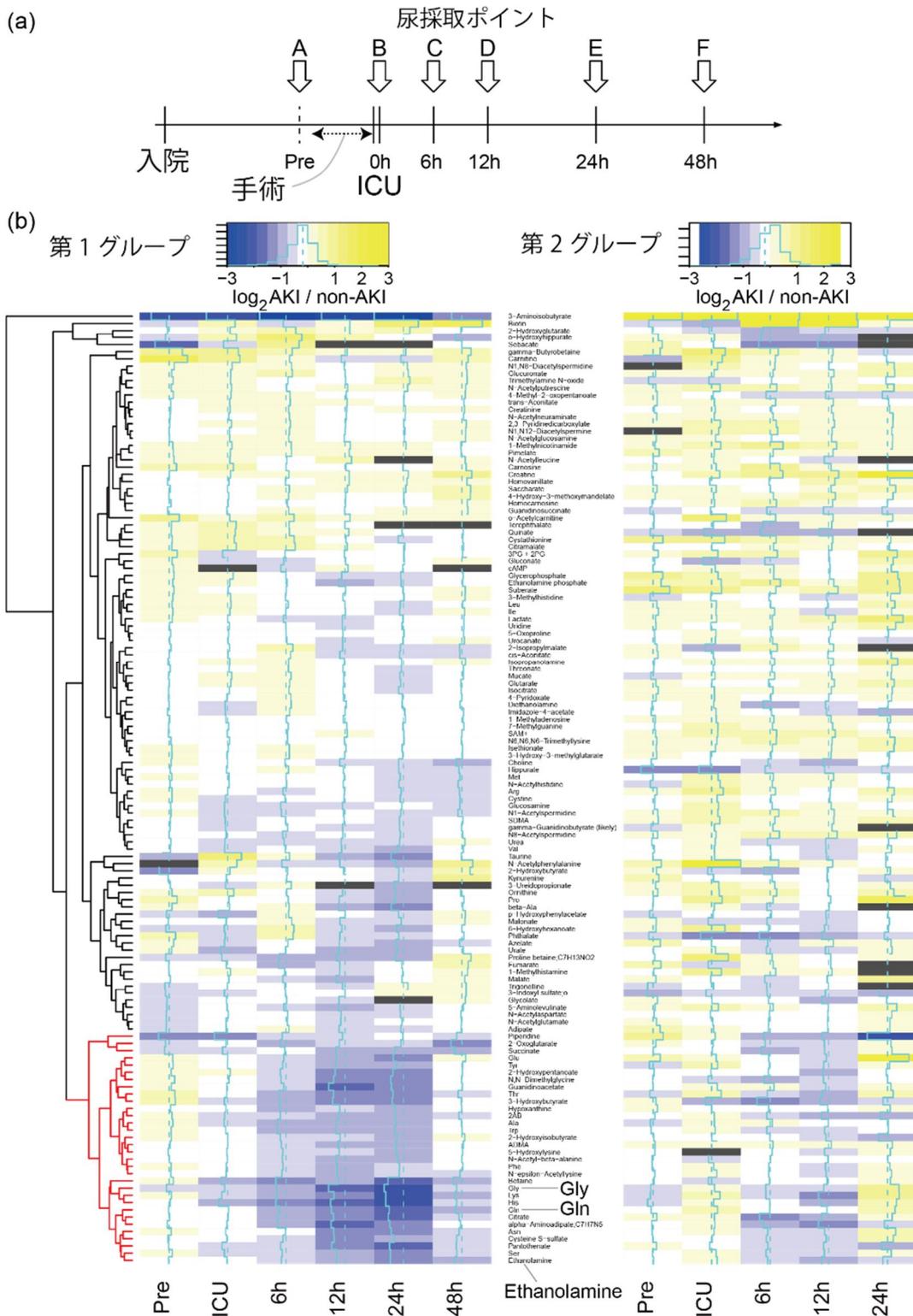


図2: 急性腎症(AKI)を発症した患者の時系列尿中メタボロームプロファイル。(a) 尿の採取ポイント。(b) AKIを発症しなかった患者(non-AKI)とAKIを発症した患者のメタボローム比較プロファイル(log₂比)。121名の患者を2回の測定(第1グループ: 61人、第2グループ: 60人)に分けて分析を行った。

そこで我々は、CE-TOFMSを用いた時に、AKIを発症した患者の尿に特有の代謝物質濃度の変化を発見できないか調べるべく、手術後に集中治療室(ICU)に入室した患者121人(その後、AKIを発症しなかった患者:63人、AKIを発症した患者:58人)の尿中メタボロームの分析を行った。尿サンプルの収集は手術前(Preと略す)、ICU入室時(ICUと略す)、および6、12、24、48時間後の6つの時点で行った(図2a)。また分析は第1グループ61人と第2グループ60人に分けて行った。

その結果、特に第1グループでAKIを発症しなかった患者、軽度のAKI発症患者、重度のAKI発症患者の間で尿中の濃度の有意差が特定の時点(12および24時間)で観測された(図2b)。例えば、エタノールアミン(Ethanolamine)とグリシン(Glycine)、グルタミン(Gln)について時系列変化を追うと、6時間からAKIを発症した患者で濃度が低下を始め、12および24時間でさらに濃度が低下していることが読み取れる。また、未知ピークAU034 ($C_9H_{16}N_2O_4$)やAU014 ($C_6H_{11}NO_4$)等についても類似した低下が観測された。

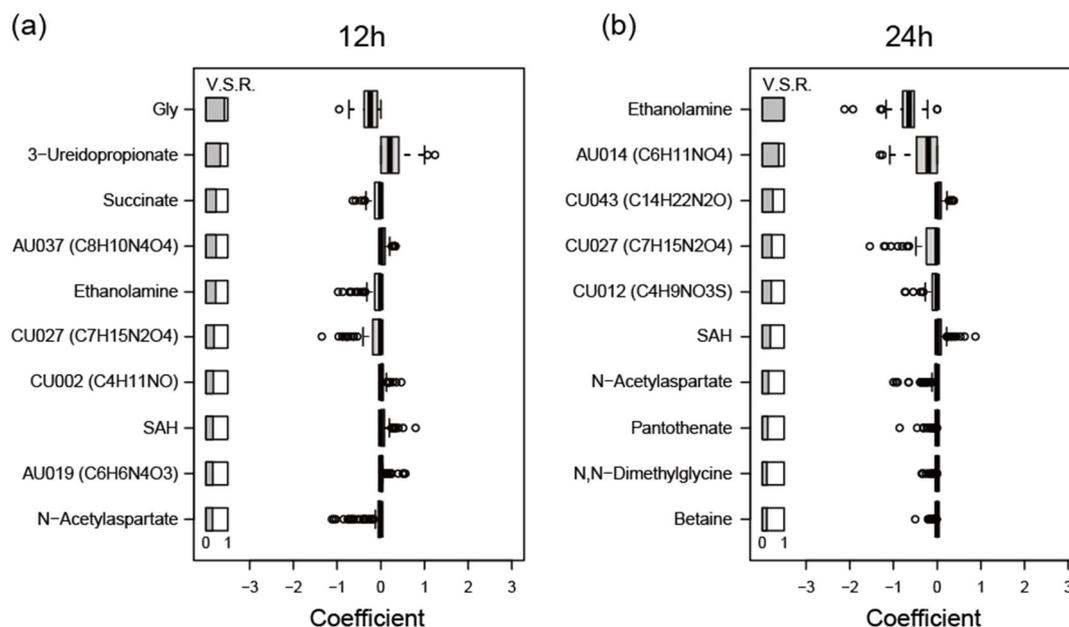


図3:急性腎症(AKI)を予測する上で、有効な代謝物質として機械学習が抽出したものの例。ICU入室から(a)12時間および(b)24時間経過した時点の尿中代謝物質を用いた結果。縦軸ラベルは代謝物質。AUまたはCUから始まるものは未知ピーク。推定される組成式が括弧の中に示されている。横軸はAKIを予測するときに各代謝物質に掛けられる係数。100回の交差検証の中で、係数がどのように分布しているかを示している。V.S.R.(variable selection rate)は、100回の交差検証の中で、各代謝物質がAKI予測に有効な変数として選択された頻度を表している。

そこで、第1グループについて、12-24時間後時点のメタボロームデータからLASSOと呼ばれる機械学習を用いてその後AKIに罹患するか否かを予測できるか、その精度検証を行った(第2グループの検証については、Saito R *et al.* *Metabolites* 2021 参照)。その結果、精度を示すAUCは12時間の時点で70%、24時間の時点では89%程度となった。同じサンプルを用いて、バイオマーカー候補として知られているタンパク質NGALの測定を行い、それを元にAKIの予測を行ったところ、AUCは12時間で55%、24時間で67%という結果となり、メタボロームデータを用いた時の予測精度がNGALのそれを上回る結果となった。機械学習がAKIの予測に有用な代謝物質として選択していたのは、予想通りエタノールアミンやグリシンだったが、AU014($C_6H_{11}NO_4$)をはじめ、いくつかの未知ピークも選ばれていた(図3)。

このように本研究によって、AKIを発症する前にどのような物質がどの時点で変化するか、またどの時点のどの物質が発症予測に有効か、その輪郭が見えてきた。

(3) まとめと今後の展望

本研究期間内に、メタボローム生データの初期の処理(プリプロセッシング)がデータの有効活用のために重要であると認識し、その方法の開発ならびに評価に力を注いだ。そしてその結果を踏まえた上で、時系列尿中メタボロームプロファイルからAIを用いてAKIの発症を予測する方法の開発を行った。一方、当初の計画の一部であった、メタボロームを制御している遺伝子群やその間の制御ネットワークの予測に関してはまだ進行途中であり、今後はそれをさらに進めていきたいと考えている。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件/うち国際共著 1件/うちオープンアクセス 2件）

1. 著者名 Saito Rintaro, Sugimoto Masahiro, Hirayama Akiyoshi, Soga Tomoyoshi, Tomita Masaru, Takebayashi Toru	4. 巻 10
2. 論文標題 Quality Assessment of Untargeted Analytical Data in a Large-Scale Metabolomic Study	5. 発行年 2021年
3. 雑誌名 Journal of Clinical Medicine	6. 最初と最後の頁 1826
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/jcm10091826	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 Saito Rintaro, Hirayama Akiyoshi, Akiba Arisa, Kamei Yushi, Kato Yuyu, Ikeda Satsuki, Kwan Brian, Pu Minya, Natarajan Loki, Shinjo Hibiki, Akiyama Shin'ichi, Tomita Masaru, Soga Tomoyoshi, Maruyama Shoichi	4. 巻 11
2. 論文標題 Urinary Metabolome Analyses of Patients with Acute Kidney Injury Using Capillary Electrophoresis-Mass Spectrometry	5. 発行年 2021年
3. 雑誌名 Metabolites	6. 最初と最後の頁 671
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/metabo11100671	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 斎藤輪太郎, 平山明由	4. 巻 47
2. 論文標題 CE-MSメタボロミクスで迫る腎臓病研究	5. 発行年 2022年
3. 雑誌名 JSBMS Letters	6. 最初と最後の頁 4-11
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 -

〔学会発表〕 計2件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 Rintaro Saito, Masahiro Sugimoto, Akiyoshi Hirayama, Tomoyoshi Soga, Masaru Tomita, Toru Takebayashi
2. 発表標題 Quality assessment of untargeted analytical data in a large-scale metabolomic study
3. 学会等名 17th Annual Conference of the Metabolomics Society METABOLOMICS 2021（国際学会）
4. 発表年 2021年

1. 発表者名 齋藤 輪太郎, 平山 明由, クワン ブライアン, プ ミニア, ナタラジャン ロキ, 新城 響, 秋山 真一, 富田 勝, 曾我 朋義, 丸山 彰一
2. 発表標題 CE-MSを用いた急性腎障害患者の尿メタボロームの解析
3. 学会等名 第21回日本再生医療学会総会 (招待講演)
4. 発表年 2022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
研究協力者	丸山 彰一 (MARUYAMA Shoichi) (10362253)	名古屋大学・医学系研究科・教授 (13901)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
米国	University of California, San Diego		