

質量分析インフォマティクス研究会・第7回ワークショップ (2022年)

## 質量分析オミクスデータを深く読む

### 開催概要

開催日時： 2022 年 4 月 22 日 (金)  
10 時 30 分 ～ 18 時 30 分 (10 時開場)  
(意見交換会 18 時 30 分 ～)

開催場所： オンライン開催 (ZOOM)

主 催： 質量分析インフォマティクス研究会※ (<http://ms-bio.info/>)  
(日本バイオインフォマティクス学会 (JSBi)) (<http://www.jsbi.org/>)



協 賛： 理化学研究所 環境資源科学研究センター (CSRS)



※本ワークショップは、質量分析インフォマティクス研究会が、JSBiの公募研究会として活動する一環として開催しています。



## プログラム

1. 開会挨拶・概要説明	10:30~10:40 (10分)
2. セッション1	
招待講演 1	10:40~11:20 (40分)
多能性維持に必要な転写後制御を受けた遺伝子群の同定	
岩崎 未央 (京都大)	
招待講演 2	11:20~12:00 (40分)
熟練作業者に匹敵する自動ピークピッキング法の開発	
金澤 慎司 (島津製作所／大阪大)	
休憩	12:00~13:30 (90分)
3. セッション2	
データ解析セッション 「プロテオミクスデータ解析」 イントロダクション	13:30~13:40 (10分)
データ解析セッション講演 1	13:40~14:00 (20分)
プロテオミクスデータの素性と扱い方 -RAWから始まるデータ解析-	
早川 英介 (沖縄科技大)	
データ解析セッション講演 2	14:00~14:20 (20分)
質量情報から配列情報へ —— どこまで信頼できるのか	
吉沢 明康 (富山国際大／京都大)	
データ解析セッション講演 3	14:20~14:40 (20分)
大規模データベースサーチの憂鬱	
三浦 信明 (新潟大)	
データ解析セッション講演 4	14:40~15:00 (20分)
2022年現状のプロテオーム定量解析手法 -single cell解析からbulk解析まで-	
岩崎 未央 (京都大)	
休憩	15:00~15:15 (15分)



#### 4. セッション3

データ解析セッション講演5 15:15~15:35 (20分)

プロテオームデータの登録・公開

奥田 修二郎 (新潟大)

プロテオミクス・セッション総合討論 15:35~15:55 (20分)

一般講演 15:55~16:05 (10分)

*In silico*解析によるRhoGAP蛋白質の相互作用ネットワーク

河内 全 (愛知県医療療育総合センター)

招待講演3 16:05~16:45 (40分)

生データの目視でメタボロミクスをするためのツールGrassHopperのご紹介

草野 博彰 (京都大)

休憩 16:45~17:00 (15分)

#### 5. セッション4

招待講演4 17:00~17:40 (40分)

KNApSack Family DB : 閉鎖まであと6年、それでも頑張るDB構築！

金谷 重彦 (奈良先端大)

6. パネルディスカッション・総合討論 17:40~18:20 (40分)

7. 研究会活動報告・閉会挨拶 18:20~18:30 (10分)

8. 意見交換会 18:30~

## 講演要旨

### 招待講演 1

#### 多能性維持に必要な転写後制御を受けた遺伝子群の同定

京都大学 iPS細胞研究所  
岩崎 未央

生体や細胞内における分子機構を解き明かすため、マイクロアレイやRNAシーケンシングといった網羅性の高い核酸解析技術を用いて、細胞内の遺伝子発現量が解析されてきた。一方、細胞内で実際に機能を発揮しているタンパク質は、ナノ液体クロマトグラフィー-質量分析 (nanoLC-MS/MS) による分析技術 (プロテオミクス) を用いて解析されてきたが、その網羅性は核酸解析技術に及ばない時代が続いてきた。しかし、近年の質量分析技術の進歩により、約1万種類のタンパク質の同定・定量が可能となってきた。

本発表では、網羅的な核酸解析技術とタンパク質解析技術による遺伝子発現量解析結果を統合し、これまで技術的に見逃されてきた転写後制御を受ける遺伝子群に着目した研究を示す。

## 招待講演 2

## 熟練作業者に匹敵する自動ピークピッキング法の開発

島津製作所

大阪大学・島津分析イノベーション協働研究所

金澤 慎司

分析サンプル中の代謝物含量の情報を得るため、クロマトグラムからピーク領域を判定するピークピッキング作業はバイオテクノロジーの基盤技術である。これまでに、自動ピークピッキング法の開発が進められているが、熟練作業者のレベルに達していないため、確認・手動修正に数時間から半日程度の時間が費やされている。近年、深層学習を用いた自動ピークピッキング法の開発により精度向上が試みられているが、未だ熟練作業者レベルには達していない。そこで、自動ピークピッキング作業の正確さを熟練作業者並みとし、修正作業を不要にすることを目的とした。そのため、ピークの正確な開始点と終了点でラベル付けされた疑似クロマトグラムを生成する新たな技術を開発し、生成した大量の疑似クロマトグラムを用いてピークピッキングニューラルネットワークの深層学習を行った。その結果、U-Net[1]がニューラルネットワークとしてSingle Shot MultiBox Detector[2]より優れていることを示した。さらに、構築したピークピッキングニューラルネットワークと熟練者として、LC-MS ワイドターゲットメタボロミクステストデータを用いて性能比較を行ったところ、両者は同等の性能を示し、ピークピッキング作業にかかる時間が実質ゼロであることが示された。

## 参考文献

- 1) Ronneberger, O., Fischer, P., and Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation, pp. 234e241, in: Navab, N., Hornegger, J., Wells, W. M. and Frangi, A. F. (Eds.), Medical Image Computing and Computer-Assisted Intervention e MICCAI 2015. Proceedings of the 18th International Conference, Munich, Germany, October 5-9, 2015, Part III. Lecture notes in computer science, vol. 9351. Springer International Publishing.
- 2) Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C., 2016. SSD: Single Shot MultiBox Detector, pp. 21e37, in: Leibe, B., Matas, J., Sebe, N. and Welling, M. (Eds.), Computer vision e ECCV 2016. Lecture notes in computer science vol. 9905.

## 招待講演 3

## 生データの目視でメタボロミクスをするためのツールGrassHopperのご紹介

京都大学 生存圏研究所  
草野 博彰

質量分析器を検出器とする液体クロマトグラフィー装置(LC-MS)は、質量電荷比と保持時間の2種類の値を利用することで検出物を区別し、シグナル強度を記録する。ワンシリーズの試料群をLC-MSで測定して得られた生データ群は、化合物組成の量的変化を示す膨大な情報を内包している。GrassHopperは、このメタボロミクスの意味を、生データの目視で直感的に把握するためのツールである。

GrassHopperでは各測定データそれぞれに色を割り当て、質量電荷比と保持時間をX軸とY軸、シグナル強度をZ軸とする3次元グラフ空間上に同時に表示する。例えば2つの実験区から得た測定データに赤と緑を割り当ててGrassHopperで表示すると、ピークの穂先の色はその化合物が多く検出された方の色で占められる。タイムコースの実験などでは、化合物量の経時的变化がピークの色グラデーションとして表されるほか、タイムラプスアニメーションとしての表現も可能である。このピーク模様の違いは、保持時間が全く同じ2種類の化合物をフラグメントイオンでなく別の化合物であると識別することなどの助けともなる。この3次元グラフ空間は自由な視点移動、回転、拡大縮小が可能で、画面上のカーソルを操作してピークを選択することで、簡単なライブラリの検索と情報の表示が可能である。また、選択したピーク群のシグナル量等の情報はファイルに出力することができるので、ExcelやR等を使った解析につなげることができる。

GrassHopperの操作はゲームパッドの使用を想定したものだが、マウス、キーボードでも操作できる。OSはWindows, Mac, Linuxで動作を確認している。分析装置、分析条件、またLC-MS装置のメーカーが異なるデータ群や、シリーズ測定期間中のメンテナンスなどでズレが生じたデータシリーズについても既知化合物の情報を利用して保持時間と質量電荷比のズレを補正して取り扱うことができる。GrassHopperのダウンロードと使用方法は以下のサイトURLで公開している\*。

\* <https://github.com/kusano-kyotouniv/GrassHopper>

## 招待講演4

### KNApSack Family DB：閉鎖まであと6年、それでも頑張るDB構築！

奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域  
金谷 重彦

オミクスと薬用/食用の知識を統合的に扱ったプラットフォームに従ってデータベースを構築すれば、社会の最重要課題である「健康」「医薬」を課題とした情報を体系的に検討できる。そこで、メタボローム研究を中心に薬用・植物知識ベース（機能性、配合）、さらにヒト生理活性を統合的に扱うデータベースKNApSack Family DB ([http://kanaya.naist.jp/KNApSack\\_Family/](http://kanaya.naist.jp/KNApSack_Family/))の構築を進めている。KNApSack Core Systemには、生物種と二次代謝物の関係データ情報が整理されており、現在までに、141,486レコードの生物種-二次代謝物の関係、二次代謝物の総数は57,906種となっている。収録されている生物種は24,690種である。白井博士（長浜バイオ大学）が開発した代謝物の三次元グラフマッチングアルゴリズム(COMPRIG)により、Twins DBにおいては二次代謝物間の類似性を検索することが可能になった。現在までに、生物種、二次代謝物にかかわる15種のデータベースの開発を進めている。本講演では、KNApSack DBにおける代謝物と生物種の関係データベースを中心に、現在構築を進めている二次代謝物合成データベースCobWeb (<http://kanaya.naist.jp/CobWeb/top.jsp>)を紹介する。また、深層学習の一つであるグラフ・コンボリューション・ネットワークにより化学構造によるアルカロイド生合成経路における開始物質の予測を行ったところ、95%程度の精度で予測が可能になった。また、現在進めている代謝経路による二次代謝物分類についても紹介したい。

## 参考文献

- 1) Afendi FM et al., KNApSack family databases: integrated metabolite-plant species databases for multifaceted plant research, *Plant Cell Physiol.* 53, e1 (2012).
- 2) Saito M, Takemura N, Shirai T\*, Classification of ligand molecules in PDB with fast heuristics graph match algorithm COMPLIG. *J. Mol. Biol.* 424, 379-390 (2012)

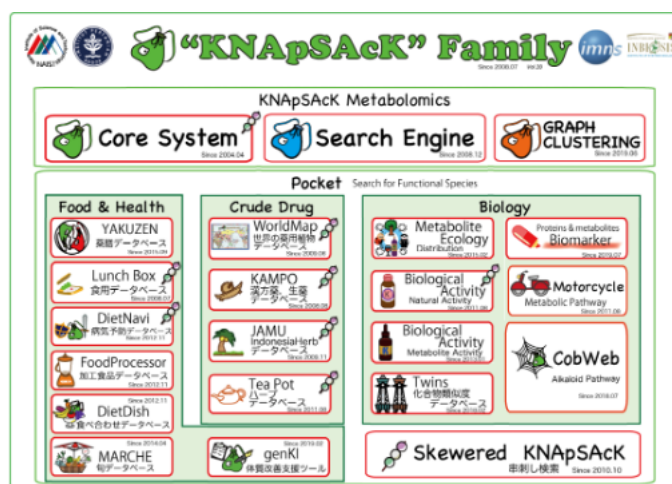


図1 KNApSack family DBのメインウィンドウ



## データ解析セッション講演1

### プロテオミクスデータの素性と扱い方 -RAWから始まるデータ解析-

沖縄科学技術大学院大学  
早川 英介

本発表では代表的な分析法であるDDA, DIAを例にとり、それぞれの長所・短所とともに、そこから生成される生 (raw) のスペクトルデータの素性に関しての解説を行う。近年のパッケージ化された解析ツールでは見落としがち・陥りがちな注意点とともに、rawデータを自分で扱う簡単なプログラミングに関する紹介も行う。





## データ解析セッション講演2

### 質量情報から配列情報へ ——どこまで信頼できるのか

富山国際大学  
京都大学 大学院薬学研究科  
吉沢 明康

ピークリストとして得られた試料の $m/z$ 情報からタンパク質を同定（アノテーション）する方法は、化学分析やメタボロミクスにおいて分子をアノテーションする方法とは、一見大きく異なっている。しかしこれらは共に、試料から得られたピークリストと標品のピークリストを比較する手法であり、タンパク質のアミノ酸配列の同定方法は、むしろ核酸の塩基配列の同定方法と大きく異なっている。

本講演では主に、DDAデータのピークリストからペプチドのアミノ酸配列を同定する手法とその代表的なアルゴリズムを紹介し、解析方法の視点から見たときに、質量分析のデータとゲノム情報がどこで接点を持つのかについて述べる。併せて、「多重検定」として知られる同定結果の信頼性の検証（False Discovery Rateによる検定）の具体的方法について紹介し、他のオミクス・データの場合と比べてプロテオーム・データでは何が異なり、何が困難であるかについて論じる。最後に、同定ペプチド情報からタンパク質を同定することの問題点についても簡単に触れる予定である。



## データ解析セッション講演3

### 大規模データベースサーチの憂鬱

新潟大学 大学院医歯学総合研究科  
三浦 信明

ゲノムが未知の生物のプロテオミクスなどでは、近縁種をかき集めて配列データベースを作る。メタプロテオミクスはその最たるもので腸内細菌などの解析を行おうとすると未知種だらけである。こうした状況下でのデータベースの構築・選択がデータベースサーチやFDR計算に及ぼす影響について考察する。



データ解析セッション講演4

2022年現状のプロテオーム定量解析手法  
-single cell解析からbulk解析まで-

京都大学 iPS細胞研究所  
岩崎 未央

本発表では、2022年における最新のプロテオーム定量解析手法についての概説を述べる。

まず、これまでよく使われていたData-dependent acquisition (DDA)手法だけではなく、Data-independent acquisition (DIA)手法における定量値の特徴と利点・欠点について示す。次に、peptideレベルの定量結果からタンパク質レベルの定量結果をどのように算出しているかを示す。最後に、近年注目されているsingle cell proteomicsでなぜTMT定量が頻繁に使われているかなど、今後のプロテオミクス解析における課題を示す。



## データ解析セッション講演5

### プロテオームデータの登録・公開

新潟大学 医学部メディカルAIセンター  
奥田 修二郎

jPOSTプロジェクト (<https://jpostdb.org/>) では質量分析に基づいたプロテオームデータの登録のためのリポジトリを開発・運用している。ここでは、jPOST repositoryの活用方法に加え、データ論文との連携について紹介する。



一般講演

*In silico* 解析による RhoGAP 蛋白質の相互作用ネットワーク

愛知県医療療育総合センター 発達障害研究所  
河内 全

RhoGAP(RhoGTPase-activating protein)は哺乳動物で約 70 種類存在し、RhoGTPase 蛋白質ファミリーの GTP 加水分解を促進して活性を制御する。RhoGAPは一般的に基質特異性が厳密ではなく、また細胞種特異的な活性調節を受けると考えられている。質量分析による相互作用蛋白質公開データを一部利用して、ヒトに特異的配列を持つ SYDE1/2 や種々の細胞内膜に局在する輸送調節因子と相互作用する ArhGAP21/23 蛋白質の基質や結合蛋白質群について予測し、情報科学を応用して構造的側面から解析した研究を紹介する。近年知的障害や自閉症で見られる点突然変異や欠失変異、コピー数多型として RhoGAP 蛋白質が同定される例が報告されており、種々のバイオインフォマティクス手法を組み合わせた解析は複雑な活性制御を受ける RhoGAP 蛋白質の生理機能理解の一助になると考えられる。



## お知らせ

本研究会では、ワークショップ等のイベント告知を研究会メーリングリストで行っています。イベントに関する案内を希望される方は是非ご登録ください。

メーリングリスト：<https://groups.google.com/g/ms-bioinfo>

参照ページ：<https://ms-bio.info/ml.html>



## 日本バイオインフォマティクス学会 入会のお誘い

日本バイオインフォマティクス学会（JSBi）は、我が国においてバイオインフォマティクスという学問分野を発展させ、その技術および関連事業の振興、並びにその教育基盤を確立するために、平成 11 年に設立されました。この学会は、バイオインフォマティクス分野最大の国際学会である ISCB（International Society for Computational Biology）の地域グループとして、またアジアの関連学会の連合体である AASBi（Association of Asian Societies for Bioinformatics）のメンバーとしても、活動しています。学会の主な活動としては、**年 1 回、年会を開催して会員の研究発表の場を提供しているほか、各種の研究会および地域部会活動、日本バイオインフォマティクス学会賞、Oxford Journals-JSBI Prize の授与、JSBi Bioinformatics Review 発行、ニュースレター発行**などを通して、多くの会員にとって意義のある学会であるように努めています。

例えば、以下のような方のご入会をお待ちしております。

- ☐ バイオインフォマティクス分野の**最先端研究に関する研究討議**を行いたい方
- ☐ バイオインフォマティクス分野での**研究交流やネットワーキング**を行いたい方
- ☐ バイオインフォマティクス分野の**知識や技術を新たに身につけたい方**
- ☐ バイオインフォマティクス分野の**共同研究者や人材を探されている方**
- ☐ バイオインフォマティクス分野に**関連する製品やサービスを開発・提供されている方**
- ☐ その他、バイオインフォマティクス分野に関する**ご関心をお持ちの方**

ぜひ多くの方にご入会いただき、今後のさらなる活動の発展にご協力いただければ幸いです。

## 入会方法のご案内

入会申し込みは学会ウェブページ（[www.jsbi.org](http://www.jsbi.org)）から随時受け付けております。入会金は1,000円、年会費は5,000円（正会員）・2,000円（学生会員）、また賛助会員は1口50,000円となります。



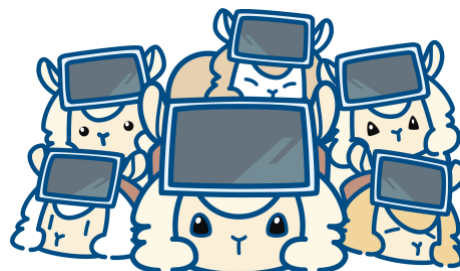
## アンケートへのご協力をお願い

このたびは、日本バイオインフォマティクス学会（JSBi）の公募研究会にご参加いただきありがとうございました。つきましては、簡単なウェブアンケート（<https://www.jsbi.org/activity/koubo/questionnaires/>）にご記入のうえ、ご意見・ご感想をお聞かせくださいますようお願いいたします。

携帯からもご回答いただけます



ウェブアンケートページ



バイオインフォひつじ（案内人）