



質量分析インフォマティクス研究会・第5回ワークショップ (2020年)

質量分析とデータ科学で生命に迫る

開催概要

開催日時： 2020年8月7日(金)

午前11時～18時30分(午前10時30分開場)

開催場所： オンライン開催 (ZOOM Webinar)

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/>)

京都大学 大学院薬学研究科

理化学研究所 環境資源科学研究センター (CSRS)

協賛： JST/NBDC統合化推進プログラム「プロテオームデータベースの機能深化と
連携基盤強化」 (<https://jpostdb.org/>)

JST/戦略的創造研究推進事業 (CREST) 『情報計測』 「質量分析と統計解析の
融合によるメタプロテオミクス」

※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

0. 開会挨拶	11:00~11:10 (10分)
開会の辞	加藤 博章 (京都大 院薬 研究科長)
概要説明	
<hr/>	
1. セッション1	
バイオインフォマティクス	11:10~11:50 (40分)
データベースから見た質量分析インフォマティクスとバイオインフォマティクス	
	五斗 進 (ROIS DS DBCLS)
<hr/>	
一般講演	11:50~12:00 (10分)
T-1 メタボロームのアノテーションは大変だ	
	荒 武 (京都大 生存圏研)
<hr/>	
休憩	12:00~13:30 (90分)
<hr/>	
2. セッション2	13:30~14:10 (40分)
代謝工学	
バイオ×デジタルの技術融合による有用微生物「スマートセル」開発への挑戦	
	蓮沼 誠久 (神戸大 先端バイオ工学セ)
<hr/>	
一般講演	14:10~14:20 (10分)
T-2 畳み込みニューラルネットワークを用いたピーク分離のパラメータ最適値の予測	
	藤田 雄一郎 (島津製作所)
<hr/>	
休憩	14:20~14:30 (10分)
<hr/>	
3. セッション3	14:30~15:10 (40分)
グライコミクス	
質量分析から見るグライコームの夢	
ーグライコミクスのインフォマティクスとマルチオミクスへの取り組み	
	三浦 信明 (新潟大 院医歯学)
<hr/>	
一般講演	15:10~15:20 (10分)
T-3 メタボロミクスの自動化を目指して ~AIを用いて開発したピークピッキング、前処理・測定メソッド・データ自動可視化のパッケージ化、解釈支援システムの開発~	
	飯田 順子 (島津製作所 分析事業部・大阪大 島津分析イノベ研)
<hr/>	
休憩	15:20~15:30 (10分)



4. セッション4	15:30~16:10 (40分)
メタボロミクス	
メタボロームを代謝指数とした表現型解析	
	福崎 英一郎 (大阪大 院工)

一般講演	16:10~16:20 (10分)
T-4 生物工学会バイオインフォマティクス相談部会における取り組み	
	堀之内 貴明 (理研 BDR)
休憩	
16:20~16:30 (10分)	
5. セッション5	16:30~17:10 (40分)
プロテオミクス	
プロテオームデータのインテリジェント解析とデータベース化	
	石濱 泰 (京都大 院薬)

一般講演	17:10~17:20 (10分)
T-5 主成分分析を用いた種々の大腸菌small ORF過剰発現時のプロテオーム変動の解析	
	丹羽 達也 (東京工大 科技創成院)
休憩	
17:20~17:30 (10分)	
6. パネルディスカッション・総合討論	17:30~18:10 (40分)
7. 研究会活動報告・閉会挨拶	18:10~18:20 (10分)
(意見交換会)	18:30~19:10 (40分)

講演要旨

バイオインフォマティクス

データベースから見た質量分析インフォマティクスと バイオインフォマティクス

情報・システム研究機構 データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター
五斗 進

シーケンシング技術の発展により、配列データやそこから得られる修飾、変異、発現などの情報が大量に蓄積され、オミクス情報から生命に迫る様々な研究が推進されている。質量分析の分野でも技術や装置が大きく発展してきており、網羅的にタンパク質（プロテオーム）、代謝産物（メタボローム）、糖鎖（グライコーム）を調べたデータも蓄積され、データベースも整備されつつある。特に、塩基配列と同様に、研究成果としての実験データの公開を公的資金の提供機関や論文の出版社から要求されるようになり、蓄積量は年々増加している。これらの質量分析データを利活用して生命に迫るためには、世界中で構築されているデータベースに蓄積されているデータを統合的に組み合わせて利用することが重要となってくる。そのためのプロジェクトもいくつか立ち上がっており、例えばプロテオームデータに関しては、世界中の関連機関がコンソーシアム ProteomeXchangeを形成して、コンソーシアムの各メンバーが収集するデータをまとめて扱えるようにしている。一方で、質量分析結果のスペクトルなどは共通基盤で扱っても、複数のサンプルのデータを組み合わせて解析するためには、サンプル間で共通に扱うためのデータ表現や実験条件などを記述したメタデータの整備も必要となる。そのようなデータの整備と共通基盤の開発は、プロテオーム、メタボローム、グライコームそれぞれのプロジェクトでも進められているが、それらの間を繋ぐ仕組みや、配列データを含む他のデータと繋ぐ仕組みなど、より広い枠組みでのデータ統合も重要である。我々のグループでは、そのための基盤技術として Resource Description Framework (RDF)によるデータ表現で関連データをグラフとしてシームレスに扱えるようにするための技術開発を進めている。本講演では、様々なデータベースの現状を概観するとともに、RDFによるデータ統合について、いくつかの事例について紹介しつつ、質量分析インフォマティクスに対してどのようなアプローチができるのか議論したい。



代謝工学

バイオ×デジタルの技術融合による 有用微生物「スマートセル」開発への挑戦

神戸大学 先端バイオ工学研究センター
蓮沼 誠久

次世代シーケンサーの開発、分析装置の高解像度・高感度化、情報解析技術の進展により、バイオ関連データが爆発的に増加している。他方、ゲノム編集やDNA合成に代表される遺伝子工学ツールに革新が起こり、バイオ操作を自動で行うラボオートメーションが実装されつつあり、ビッグデータから抽出した有用情報を活用し、先端的な遺伝子工学で生物代謝を改変する Engineering Biologyの社会実装が世界的に進んでいる。我が国ではサイバーとフィジカルが高度に融合した「超スマート社会」を未来の姿として共有し、これを世界に先駆けて実現するための取り組みを強力に推進している。バイオモノづくり分野では爆発的に増加するバイオデータとバイオテクノロジーを有効に繋ぐことにより、化学品、食品、その他の新機能材の創出がこれまでにない期間、コスト、性能で開発できることが期待されている。こうした背景の下、「高度に機能がデザインされ、機能の発現が制御された生物細胞」をスマートセルと定義し、スマートセルを用いた次世代産業「スマートセルインダストリー」の構築が期待されている。

微生物は代謝系の中で多くの化合物を作り出すことができることから、その機能を活用することで、高効率な物質製造プロセスを構築することができる。しかしながら、細胞の代謝システムは複雑に制御されており、従来の手法では生産性の向上に限界があった。そこで我々は、バイオインフォマティクスによる代謝経路設計技術や長鎖DNA合成技術、動的メタボロミクス等の要素技術を開発し、これらを有機的に連携させた『スマートセル創出プラットフォーム』の開発を行うことで、微生物の物質生産能力を最大限引き出す先端バイオ工学の構築に取り組んできた。

情報解析技術を有効に活用するためには、大規模で多様性に富んだデータセットを体系的に取得する必要がある。我々は遺伝子操作と代謝変動の因果関係を解析する手段として、細胞内代謝物網羅的解析技術である「メタボロミクス」を重要視してきたが、メタボロミクスにはスループット性と再現性の課題があった。これまでの研究からこの課題の大部分は「リアクター内の細胞から代謝物を抽出するまでの前処理工程」に由来していることが分かっている。そこで本研究では



前処理工程を自動化したロボティクスの開発を行った。本講演では、ロボティクスの開発によりハイスループット化・高精度化したメタボロミクスを紹介したい。

また、計算科学と合成生物学の融合により構成される「設計 (Design) -構築 (Build) -評価 (Test) -学習 (Learn) のDBTLサイクル」により、バイオ生産の向上に成功したベンジルイソキノリンアルカロイド (BIA) 生産菌の開発にも触れ、スマートセル創出プラットフォームの現状について紹介したい。



グライコミクス

質量分析から見るグライコームの夢 ーグライコミクスのインフォマティクスとマルチオミックスへの取り組み

新潟大学 大学院医歯学総合研究科
三浦 信明

ガラクトースやマンノース, N-アセチルグルコサミンなど糖分子が鎖状に連なった糖鎖は, 分子認識, タンパク質分子の品質管理, 細胞内外でのシグナル伝達などを担っている. そして, 細胞分化や疾患をはじめとしたほぼすべての生命現象にその姿を現す. さながら細胞や生体分子における顔であり鍵でありタグとして働く. その全カタログを解明する研究がグライコミクスである. 糖鎖には遊離の糖鎖以外に, タンパク質に結合するN-結合型およびO-結合型糖鎖, 脂質に結合する糖脂質糖鎖, プロテオグリカン形成するグリコサミノグリカンなど多くの種類があり, それぞれ化学的な環境が異なるため分析法も異なる. その上, 確立した簡便なプロトコルが無い. 重要と言われ続けているが置いてきぼりにされてきた原因の一端はここにある.

講演者の共同研究者らは, 糖鎖の網羅的解析を目指して様々なサンプルの網羅的解析によるグライコミクスを進めつつ, 安定・簡便なプロトコル, より多くの糖鎖, 結合様式等の帰属可能な方法などの開発を勢力的に進めている[1]. 講演者は彼らと共同で, 膨大な質量分析データから糖鎖の帰属, 発現量の変動などの生合成経路上での表現に取り組んできた. 現在まさに発展をしている分野であるが故に情報処理ツールが少ない事が課題であり, これらグライコミクスをサポートするソフトウェアをToolbox accelerating glycomics (TAG)という名称で開発している.

グライコミクスによって糖鎖を帰属し, 疾患やその他生命現象による発現量の変動を追えるようになったとして, その糖鎖がどのような働きをしているか, その糖鎖はなぜ増えるのかなどを理解するのは大仕事である. 糖鎖発現, 生合成経路からマルチオミックス分析にもつながるようなツールが必要である. この点に関して, 糖鎖を入り口として疾患やマルチオミックスを俯瞰できるようなデータベースGlyco-Net[2]を開発した.

講演者自身は異分野からの糖鎖研究への参入者である. 上述のような取り組みの一端をご紹介することで, 「それなら自分にもできる」といった参入者が増えることを望んでいる. より良いツールができることによって, 糖鎖研究が加速され, 最終的に生命科学研究が発展すると思っている. そうなれば良いと思っている.



- [1] Shinohara Y and Furukawa J-I, 化学と生物, 53, 586-592 (2015),
Fujitani N et al, Proc Natl Acad Sci USA, 110, 2105-2110, (2013) DOI:10.1073/pnas.1214233110
- [2] Hashimoto R et al. BMC Systems Biology 2010, 4:91 (2010), DOI: 10.1186/1752-0509-4-91,
Miura N et al, Glycoinformatics, Methods in Molecular Biology 1273, Lütteke T and Frank M
(eds.), Springer, 149-157 (2015), DOI: 10.1007/978-1-4939-2343-4_11



質量分析インフォマティクス研究会
(日本バイオインフォマティクス学会)



メタボロミクス

メタボロームを代謝指紋とした表現型解析

大阪大学 大学院工学研究科
福崎 英一郎

TBC



プロテオミクス

プロテオームデータのインテリジェント解析とデータベース化

京都大学 大学院薬学研究科
石濱 泰

日本内外に散在している種々のプロテオーム情報を標準化・統合・一元管理し、多彩な生物種・翻訳後修飾・絶対発現量も含めた横断的統合プロテオームデータベースを開発すべくjPOST (Japan ProteOme STandard Repository/Database) プロジェクトは2015年に産声を上げた。すでに確立した国際標準リポジトリおよび他のプロテオームデータ基盤にはない精度の高いデータ標準化機能を有しており、アジアオセアニア地域にとどまらず、世界中より幅広いプロテオームデータの受け皿として順調に発展している。本発表では、現状の問題点や今後の方向性について議論したい。更に、高度な階層性を有するプロテオームLCMSデータに対し、最先端の統計的手法を駆使した新しい解析法についても紹介する。



一般演題

メタボロームのアノテーションは大変だ

京都大学 生存圏研究所 森林代謝機能化学分野
荒 武

メタボローム研究でのアノテーションはとても大変です。これまでいろいろなツールやデータベースを開発してきましたが、知識ベースでこつこつアノテーションをできる人は少なく、まだほとんど手付かずの状態です。メタボロームデータのフォーマットの多様さ、データベースの未整備（ほしいデータを集めるのも一苦勞）、化合物分類の難しさ（分野や視点でかなり違うし、オントロジーもまだまだ使いにくい）、簡単な手順書などの不足、など様々な問題があります。また、ゲノム研究と同様、アノテーションしたからといって論文にはならないということも参入の大きな障壁になっています。これまで10年以上、メタボロームデータのアノテーションを行ってきましたが、かなり趣味と信念に基づいている感があります。この状況を打破し、遺伝子アノテーションなみに人々が参入してくれるようにするためにはどうすればよいのか、ざっくばらんな議論ができればと思って参加します。



一般演題

畳み込みニューラルネットワークを用いた
ピーク分離のパラメータ最適値の予測

1) 島津製作所

2) 大阪大学 島津分析イノベーション協働研究所

藤田 雄一郎¹, 野田 陽¹, 山田 洋平¹, 種田 克行¹, 飯田 順子^{1,2}, 梶原 茂樹¹

質量分析データ解析のための有用なソフトウェアが多く利用できるようになってきた。しかしながらこれらのソフトウェアの多くに共通する課題がパラメータの最適値の設定である。技術的な知識やノウハウを駆使して試行錯誤により決定する必要がある。

本研究ではピーク分離用ソフトウェアであるAMDIS (<http://www.amdis.net/>) を例にとり、畳み込みニューラルネットワーク(CNN)を用いてAMDISソフトウェアのパラメータの最適値をdata dependentに予測する手法を開発した。評価には生体メタボロームを測定したGC/MSデータを用い、AMDISのデフォルト設定を使うよりも多くのデータに対して最適な結果が得られることを確認した。



一般演題

メタボロミクスの自動化を目指して ～AIを用いて開発したピークピッキング、前処理・測定メソッド・データ 自動可視化のパッケージ化、解釈支援システムの開発～

- 1) 島津製作所 分析計測事業部
- 2) 大阪大学 島津分析イノベーション協働研究所
飯田 順子^{1,2}

研究ツールとして定着してきたメタボロミクスは、質量分析計（MS）が広く使用される。求める良好な解析結果を得るためには、堅牢で再現性の良い前処理や分析のメソッド開発が重要なポイントの一つであるが、時間がかかり経験も要求される。また、MSの測定速度が早くなり感度が高くなる結果、生産されるデータ量は膨大で、ピークピッキングから始まり解析結果を得るまでに多大な労力と時間を要している。

研究者がMSを使う目的である「新規発見」のためのデータ解釈に時間を使えることを目指し、AIを用いて開発したピークピッキングアルゴリズム、前処理・測定メソッド・データ自動可視化のパッケージ化、解釈支援のためのツールの研究開発と社会実装を進めている。この取り組みについて紹介し、ニーズを伺い、議論のスタートとしたい。

なお、AIを用いるピークピッキングは、富士通株式会社、株式会社富士通研究所との共同研究の成果である。ここへのニーズインプットを含め一連の研究開発は、大阪大学 福崎英一郎教授、松田史生教授との協業で進めている。



一般演題

生物工学会バイオインフォマティクス相談部会における取り組み

理化学研究所 生命機能科学研究センター
堀之内 貴明

名古屋大学大学院 生命農学研究科
兒島 孝明

名古屋大学大学院 創薬科学研究科
蟹江 慧

分析技術の高度化と計算機性能の向上に伴い、バイオインフォマティクスの重要性が増している。しかしながら、その実践には高度な専門性を要する点や、目的によって多種多様な方法論が存在する点など、初学者にとっての敷居は高い。また実験系研究者にとって、インフォマティクスの専門家との意思疎通や連携の際には様々な障壁が存在する。そこで我々は、異分野交流の場の提供や研究活動支援、若手人材育成などを目的とし、生物工学会バイオインフォマティクス相談部会（以下当部会）を2017年に設立した。毎年講演会やハンズオンセミナー等の開催、オンライン相談窓口の運営などを行っている。当部会の特徴として、実験生物学のバックグラウンドを有し、かつバイオインフォマティクスに精通した研究者が多く活動に参画することにより、初学者が直面する様々な障壁の解消を試みている。本部会は生物工学会の研究部会制度を利用して運用されているものの、こうした活動は学会内に限定されるべきではないと考え、当部会が主催する講演会等では学会外からも積極的に演者を招聘し、人材交流の機会を提供している。本ワークショップでは話題提供として当部会の活動内容について紹介し、ウェットとドライの連携の在り方などについて議論したい。



一般演題

主成分分析を用いた
種々の大腸菌small ORF過剰発現時のプロテオーム変動の解析

東京工業大学 科学技術創成研究院 細胞制御工学研究センター
丹羽 達也, 茶谷 悠平, 田口 英樹

Ribosome Profiling法の発展などにより、small ORFと呼ばれる短いORF領域が原核、真核生物を問わず数多く存在することが近年明らかになってきている。その多くは下流や近傍の遺伝子の発現制御に使われると考えられているが、その制御機構は様々であり、実際の制御機構や、その役割すらわかっていないというものも数多くみられる。そこで私たちは大腸菌が持つ種々のsmall ORFについて、過剰発現させた際のプロテオーム変動などの表現系をまとめて調べることで、それらの役割や機能などを分類できないかと考えた。独自に約30種類のsmall ORFを選抜し、それぞれを過剰発現させた際のプロテオーム変動を網羅的に数値化し、得られたデータを主成分分析によって解析した。その結果、small ORF過剰発現時のプロテオームの変動パターンは一様ではなく、大きく2つのパターンに分けられることが示唆された。