

質量分析インフォマティクス研究会・第4回ワークショップ (2019年)

データ科学が拡大する質量分析の地平線

開催概要

開催日時： 2019年3月19日(火)

午前10時30分～18時30分(午前10時開場)

開催場所： JST東京本部別館(K's五番町ビル)1階ホール

(東京都千代田区五番町7 K's五番町)

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： JST/NBDC 統合化推進プログラム「エピゲノミクス統合データベースの
開発と機能拡充」(<https://chip-atlas.org/>)

※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	10:30～10:40 (10分)
2. メタボロミクス 質量分析を用いた植物メタボロミクスデータの情報解析と共有	10:40～11:20 (40分)
	福島 敦史 (理研・CSRS)
3. 代謝工学 有用バルクケミカル合成大腸菌プラットフォーム開発における質量分析機器の活用	11:20～12:00 (40分)
	野田 修平 (理研・CSRS)
休憩	12:00～13:30 (90分)

4. ケモインフォマティクス	13:30~14:10 (40分)
複雑骨格物質の生合成単位予測	小寺 正明 (東京大・院・工)
5. 一般演題	14:10~15:00 (50分)
5-1. MS と NMR 分析から得られる漢方方剤ケミカルフィンガープリントを採り入れた漢方処方システム基礎理論の包括的解析	岡田 岳人 (徳島文理大・香川薬)
5-2. 質量分析を用いた抗体医薬の開発	安永 正浩 (国立がん研)
5-3. 代謝物アノテーションのための false discovery rate (FDR) 算出方法	津川 裕司 (理研・CSRS)
5-4. 糖鎖・糖タンパク質用質量分析リポジトリ GlycoPOST の公開	木下 聖子 (創価大・理工)
5-5. 質量分析ハッカソンへ参加して	土屋 大樹 (東京電機大・院・理工)
休憩	15:00~15:25 (25分)
6. グライコプロテオミクス	15:25~16:05 (40分)
グライコプロテオミクスにより明らかになった Notch1 受容体上の O-GlcNAc glycan の糖鎖構造多様性	小川 光貴, 岡島 徹也 (名古屋大・院・医)
7. プロテオミクス	16:05~16:45 (40分)
公共データベースを利用したリン酸化シグナル伝達経路の研究	吉崎 尚良 (金沢医科大・医)
休憩	16:45~17:10 (25分)
8. ChIP-Seq	17:10~17:50 (40分)
公共 ChIP-seq データをフル活用し、遺伝性疾患の解明や創薬に挑む	沖 真弥 (九州大・院・医)
9. パネルディスカッション・総合討論	17:50~18:20 (30分)
10. 研究会活動報告・閉会挨拶	18:20~18:30 (10分)



講演要旨

メタボロミクス

質量分析を用いた植物メタボロミクスデータの情報解析と共有

理化学研究所 環境資源科学研究センター
福島 敦史

植物は進化の過程で、多種多様な代謝物を生合成する代謝経路を発達させ、与えられた生育環境に適応してきた。近年の質量分析計や高速シーケンサに代表される計測技術の革新は、植物の生長と発達メカニズムの背後にあるゲノム、トランスクリプトーム、プロテオームおよび代謝物情報（メタボローム）等に関して多くの情報をもたらしている。講演者は、このような網羅的な分子情報を組み合わせ、植物の生長と発達の理解を目指して研究を進めてきた。本講演では、はじめに植物科学における代謝物プロファイリングおよび情報解析の現状について概観する。次に、これまで行った講演者らの植物メタボロミクス研究を通して、その背景、データ解析の実際やデータ解釈を改善する取り組みの一端を紹介する。また、JST/NBDCの枠組みの中でMetaboBank（旧MassBank）プロジェクトが開始され（2018年度～）、遺伝研を中心にメタボロームデータリポジトリの開発が進められている。我々の研究チームでは、蓄積したメタボロームデータの公開†と再解析とによって、研究の再現、既存アノテーションツールの検証、データ測定の品質向上、機械学習精度の向上等を目指している。その現状と今後についても述べる。

† <http://metabobank.riken.jp/>



代謝工学

有用バルクケミカル合成大腸菌プラットフォーム開発における質量分析機器の活用

理化学研究所 環境資源科学研究センター
野田 修平

近年、地球温暖化や化石資源の枯渇への対策など、持続可能な社会の構築に向けて取り組むべき様々な事柄が社会問題となっている。2015年にCOP21で締結されたパリ協定では、2030年度までに空気中の二酸化炭素の26%（2013年度比）を削減することが求められており、低炭素循環型社会の構築はこれまで以上にその必要性が求められている。化石資源への依存から脱し、環境負荷の少ない社会を構築するため、バイオマスや二酸化炭素などの再生可能資源を原料に用いた微生物プロセスの開発は重点的に取り組むべき研究分野であるといえる。ブタノール等の燃料生産、生分解性プラスチック原料生産に関する研究が行われている一方、発表者らのグループは現状その生産を石油資源に依存している芳香族化合物やその誘導体の生産に注目した。そして、様々な芳香族化合物やその誘導体を高生産量、高収率で生産可能な汎用性のあるプラットフォーム株の創製を試みた。

微生物で芳香族化合物を生産する際、いくつかの芳香族化合物が生成する最終反応においてピルビン酸の放出を伴う。本プラットフォームにおいては大腸菌を宿主に用い、芳香族化合物生産の最終反応においてピルビン酸を放出するという特徴を最大限生かすための代謝改変を行った。7種類の化合物において、単純に経路を導入した株と比較して生産量、収率をそれぞれ5～50倍程度、大幅に引き上げることに成功した(1),(2)。また、このプラットフォームで生産可能な化合物である3-ヒドロキシ安息香酸を起点とし、放線菌の芳香環形成経路と芳香族化合物分解菌の芳香環分解経路を組み合わせることにより、これまで微生物を用いての生産報告の例がないマレイン酸の微生物生産に成功した(3)。今回の発表では、バルクケミカルを高生産する大腸菌を創製する上での質量分析機器の活用に関しても一部言及する。

(1) Fujiwara et al. ACS Synth Biol, 2018, 7: 2698-2705.

(2) Noda S et al. Metab Eng, 2016, 33, 119-129.

(3) Noda S et al. Nat Commun, 2017, 8(1): 1153-1165.

ケモインフォマティクス

複雑骨格物質の生合成単位予測

東京大学大学院 工学系研究科
小寺 正明

抗生物質や色素など、生命維持には直接関わらない種特異的な物質を二次代謝産物という。医薬品や農業などの様々な産業分野への応用を目指し、生物・薬理活性を示す有用な二次代謝産物の生合成機構の解明に向けた研究が多く進められてきた。しかし、化学構造は明らかでも代謝経路が明らかでない化合物の数は非常に多いことから、情報科学を用いた反応経路推定手法の開発が望まれている。

生物の体内では、限られた数の生合成単位を原料として、多くの二次代謝産物が産生されている。複雑骨格を持つ機能分子の生合成機構を推定するには、その構造を生合成単位まで分解することが重要となる。そこで、本研究では、二次代謝産物の生合成単位を自動的に予測し、出発物質の同定や生合成経路推定の支援を行うプログラムの開発を目的とした。

本研究では、生合成単位を概念を拡張し、本来の意味のタイプA、代謝経路上の出発物質や重要な中間体であるタイプB、タイプBを構造変換したタイプCの三つのタイプに分類した。タイプBは二次代謝産物の生合成経路から収集され、タイプC（タイプA含む）はタイプBに構造変換ルールを適用して生成した。合計765種類の生合成単位がライブラリに収録された。

次に、クエリ分子に対する生合成単位の組合せを自動推定するアルゴリズム（以下、提案手法）を検討した。処理の流れを次に示す。生成したコンパクトな生合成単位ライブラリを利用して、クエリ分子を次々と分解し、クエリ分子から生合成単位までを結ぶネットワークを生成する。そこから得られるフラグメント間の関係を利用して、クエリ分子中の生合成単位の組合せ候補を生成する。最後に、クエリ分子を生合成単位ごとに色分けした化学構造を出力する。

KEGG PATHWAY 1.10クラスにある28代謝マップ中の943化合物を対象として、予測結果の正誤を判定した。その結果、全体の約95.8%の化合物に対して正しく予測することができた。また、約92.5%の化合物に対しては1分以内に計算が終了し、組合せ爆発を回避できたことが示された。

予測に失敗した原因の約半数は部分構造検索に関するものであった。提案手法で用いた部分構造検索手法では、芳香族性の結合と通常の単結合または二重結合が区別された。その課題を解決するためには、部分構造検索手法を新規に開発する必要がある。

今後の展望として、生合成経路未知の二次代謝産物を生合成単位まで分解した後、各単位がどの順番で結合するか予測することで、生合成経路を予測できると期待される。



グライコプロテオミクス

グライコプロテオミクスにより明らかになった Notch1 受容体上の O-GlcNAc glycan の糖鎖構造多様性

名古屋大学大学院 医学系研究科
小川 光貴, 岡島 徹也

Notch1 EGF リピートには、EGF ドメイン特異的な O-GlcNAc 転移酵素 (EOGT)によって O-GlcNAc が転移される。これまでは、O-GlcNAc の検出には 抗 O-GlcNAc 抗体やレクチン、クリックケミストリーといったツールで使用していたが、Notch1 受容体には複数の O-GlcNAc 修飾サイトが存在することが示唆されていた為に、それぞれの O-GlcNAc 修飾サイトにおける O-GlcNAc レベルを定量することは不可能であった。

そこで私たちのグループでは、質量分析装置 (Obritrapp Fusion) と糖鎖解析ソフトウェア (Byonic) を用いたグライコプロテオミクスで Notch1 受容体上の O-GlcNAc の定量を試みた。その結果、特定の EGF ドメインで O-GlcNAc-Gal-NueAc 3糖構造が生じていることを見出した。また、各 EGF ドメインで O-GlcNAc 修飾レベルや O-GlcNAc glycan 構造が異なることが分かった。本ワークショップでは、私たちが実施した条件検討の部分も含めて、実際のデータや測定条件等も紹介したい。



プロテオミクス

公共データベースを利用したリン酸化シグナル伝達経路の研究

金沢医科大学 医学部 小児外科学
吉崎 尚良

タンパク質リン酸化は、最もポピュラーな翻訳後修飾で、生命現象や病態に大きな影響を持つことが知られている。近年の質量分析技術の向上は、リン酸化プロテオミクスを発展させ、大規模なリン酸化サイトの同定と、膨大な量のタンパク質リン酸化部位のデータベース化につながった。しかし、これらの膨大なリン酸化サイトの多くは生理的意義が判明しておらず、逆にその数が生理的に意義のあるリン酸化サイトの決定を困難なものにしている。我々は、ヒトゲノム上のリン酸化サイトを周辺配列の構造から、178のリン酸化モチーフに分類したデータを使い、リン酸化モチーフの進化的保存性(Conservation index:CI)がその生理機能への重要性和関係があることを見出した。がんにおけるドライバー遺伝子の多くはタンパク質リン酸化に関係する遺伝子であり、その基質であるリン酸化サイトに入る変異はがん化への影響を持つことが期待された。そこでリン酸化モチーフ情報が、疾患シグナルの探索に有効か検証もかねてリン酸化モチーフ情報と、International Cancer Genome Consortium (ICGC)(<https://icgc.org/>)に登録されたがん特異的変異データを組み合わせてその関係を調べた。がん組織では、生理的重要性の高いリン酸化モチーフに変異が蓄積しやすい傾向がみられ、この結果は、がん組織のリン酸化モチーフ上の変異が、ゲノム不安定性によるランダムな変異の挿入でなく自然選択的に挿入されていることを示唆した。このように本ワークショップでは、ウェット研究者である筆者が、慣れない公共データベースを使った解析を行い、いかに生理的に重要そうなシグナル伝達経路の探索に右往左往したか、リン酸化モチーフ上のがん突然変異の解析を例にその顛末を紹介する。



ChIP-Seq

公共 ChIP-seq データをフル活用し、遺伝性疾患の解明や創薬に挑む

九州大学大学院 医学研究科
沖 真弥

ゲノムに結合する様々なタンパク質の結合や分布を調べるため、これまでに10万件近くのChIP-seqデータが報告されている。しかしそれらを活用するためには非常に複雑で大規模な計算処理が必要なため、その多くが活用されずに死蔵されているのが現状である。そこで我々は既報のChIP-seqデータを網羅的に収集、計算、統合し、その解析結果をウェブサービスとして公開している(ChIP-Atlas; <https://chip-atlas.org>)。これにより、興味のあるゲノム領域における転写因子や修飾ヒストンの分布が視覚的に理解できるため、遺伝子制御ネットワークの解明や、エンハンサー領域の同定に応用できる。我々はこの膨大なデータを利活用し、各種組織で特異的に発現する遺伝子座やnon-coding GWAS SNPの周辺領域において、結合がenrichするような転写因子を複数同定した。また、薬剤の作用点となるような転写因子を同定おり、その結果についても紹介したい。



一般演題

MS と NMR 分析から得られる漢方方剤ケミカルフィンガープリントを採り入れた
漢方処方システム基礎理論の包括的解析

徳島文理大学 香川薬学部
岡田 岳人

漢方方剤（漢方薬）は、様々な種類の生薬を組み合わせて製される。そして漢方では、現代医学（西洋医学）における薬物療法アプローチとは異なり、患者の状態を「証」という診断基準（例：虚/実の体質）から捉え、個人の「証」に適した方剤を処方する。こうした「証」への方剤適用は、多様な治療経験・エビデンスに基づいたものであり、処方システムの全体像は複雑である。

発表者らは、この複雑な漢方処方システム基礎理論の包括的解明を目指し、方剤のノンターゲット化学分析、方剤・処方に関する研究用データベース構築、そして各種データの多変量解析を主体とするケモメトリクス研究を進めてきた [例: 1, 2]。本講演では特に、MSとNMRを用いた分析から得られた方剤構成生薬の含有成分に由来するケミカルフィンガープリントと「証」との相関を解析した結果について紹介する。

[1] Okada et al (2016) J Nat Med, 70:107—114

[2] 岡田ら 日本生薬学会 第65回年会（広島）. 2018年9月16—17日. 2B-09



一般演題

質量分析を用いた抗体医薬の開発

国立がん研究センター 先端医療開発センター
安永 正浩

理化学研究所 生命機能科学研究センター
川井隆之, 清水義宏

次世代抗体医薬として抗体薬物複合体(Antibody-drug conjugate, ADC)の研究開発を行っている。ADCは抗体にリンカーを介して抗がん剤を付加したものである。ADCは、抗体のもつパッシブターゲティングとアクティブターゲティングにより、腫瘍選択的に集積することができる。さらに、ADCにおけるリンカーテクノロジーはコンロールド・リリースとして、腫瘍或いはがん細胞内で抗がん剤を選択的に放出することができる。ADCのもつユニークな薬理効果・薬効性をいかに正確に評価できるかということが重要な課題になっている。そこで、コンロールド・リリースの評価に質量分析イメージングを活用している我々の取り組みを説明する。さらに新規抗体開発のために行っている1細胞レベルの質量分析システムについても紹介したい。



一般演題

代謝物アノテーションのための false discovery rate (FDR) 算出方法

理化学研究所 環境資源科学研究センター
津川 裕司

液体クロマトグラフィータンデム型質量分析 (LC-MS/MS) を用いたノンターゲットメタボローム解析は、現在最も使用されている分析技術である。また、解析ツールも充実してきており、現在では特別な知識が無くとも代謝研究ができるようになりつつある。しかしながら、論文などに掲載している代謝物のアノテーション情報が本当に確かなのか？最近、「我々が開発した手法により数百〜数千もの代謝物がアノテーションできます」という主張の論文が多く見受けられるようになってきたが、中身を読むとその主張がmisleadingであることが多い。簡単にメタボローム解析ができるようになってきた一方で、その結果の信頼性を担保するための手法開発が、いま必要性を増してきているのは間違いない。本発表では、このような問題を解決するための手法の一つであるfalse discovery rate (FDR) の計算方法の、メタボローム解析研究における動向を紹介する。



一般演題

糖鎖・糖タンパク質用質量分析リポジトリ GlycoPOST の公開

創価大学 理工学部
木下 聖子

新潟大学大学院 医歯学総合研究科
奥田 修二郎, 渡辺 由

糖鎖統合データベースプロジェクトでは複合糖質および糖鎖の質量分析の生データのためのリポジトリの開発を進めており、その中でグライコミクスおよびグライコプロテオミクスの質量分析データの受け入れを担うシステムとしてGlycoPOSTを開発中である。

GlycoPOSTは、すでに稼働中のプロテオームデータリポジトリシステムjPOST repositoryの機能を継承し、新規にグライコプロテオミクスのデータに特化する形で開発中であり、相互にデータの参照が可能となる他、高速なアップロード、入力作業の簡便性といった特徴も引き継がれる予定である。また、GlycoPOSTは海外のグライコミクスリポジトリUniCarb-DRとも連携しており、海外からの関心も高い。特に、MIRAGEガイドラインとの互換性があるため、MIRAGE Projectの提唱する糖鎖関連実験を報告する際のガイドラインに準拠したメタデータの登録、インポートおよびエクスポートが可能である。本発表ではGlycoPOSTの概要とその特徴、期待される役割などについて紹介する。



一般演題

質量分析ハッカソンへ参加して

東京電機大学大学院 理工学研究科
土屋 大樹

私は第2回質量分析インフォマティクス・ハッカソンに参加しました。このハッカソンには、幅広い分野の専門家が参加されており、ご意見をお伺いしながら開発を行うことで、自分一人では解決できない問題を解決し、大変有意義な時間を過ごすことができました。

また、開催期間は5日間とかなり長かったため、開発に限らず様々な分野の研究に関するお話を直接お伺いするという貴重な体験をさせていただき、大いに刺激を受けました。特にDB集約に関するお話には興味を惹かれ、自分でも何か貢献できることはないかと考えるようになりました。

今後も本ハッカソンのようなイベントがあれば積極的に参加していきたいと考えており、将来的には開催のお手伝いなどもさせていただきたいと考えています。