

構造活性相関部会・ニュースレター <1 October, 2016>

SAR News No.31

「目次」

///// Perspective/Retrospective /////		
メタボロミクス技術開発と高解像度表現型解析への応用		
傳寶 雄大、福崎英一郎	•••	1
///// Cutting Edge /////		
植物におけるオミックス統合ネットワークシステム生物学		
福島敦史	•••	9
オミクス統合解析を用いた植物二次代謝機能ゲノミクスについて		
· · · · · · · · · · · · · · · · · · ·	•••	17
///// Activities /////		
<報告>		
構造活性フォーラム 2016		
「分子標的薬の創生やゲノム創薬における目のつけどころ」開催報告	•••	26
<会告>		
第 44 回構造活性相関シンポジウム:第 31 回農薬デザイン研究会共同開催		
	•••	28
The Third International Symposium for Medicinal Sciences		30

メタボロミクス技術開発と高解像度表現型解析への応用

大阪大学大学院工学研究科 傳寶 雄大·福崎英一郎

1. はじめに

メタボロミクスは、細胞内の複雑な酵素反応の組み合わせによって生じた代謝産物総体を解析するポストゲノム研究領域の一つである。メタボロームはゲノム情報の実行課程であるセントラルドグマの最下流に位置しているため、メタボロミクスはゲノム情報を反映した有効な表現型解析手段の一つであると言える(図 1)。その応用範囲はポストゲノム科学としての微生物、植物、医学研究にとどまらず、医療診断や食品の産地判別、品質評価などと多岐にわたる。多くの代謝物は生物間での互換性がある一方、代謝物はおのおの多岐にわたる化学的性質を有し、それぞれ存在量も大きく異なるため、これらを一斉に、且つ正確に定量する標準化された技術は未だに開発途上である。また、得られた膨大な代謝物データから多変量解析などを用いて有用な変化のみを抽出する作業、すなわち「データマイニング」についても王道は存在しない。したがって、研究対象や研究目的に応じて様々な方法論や化合物ライブラリが乱立しているのが現状であり、これがメタボロミクス技術の普及を妨げる一因となっている。本稿では、メタボロミクスの技術開発と応用、これからの展望を俯瞰的に紹介する。

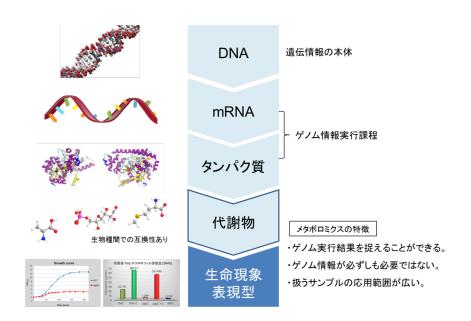


図 1. ポストゲノム科学におけるメタボロミクスの特徴

2. メタボロミクスの現状と技術開発

メタボロミクス研究の大まかな流れとしては、(1) 研究対象となる生体試料の調達などの生命科学、(2) 代謝産物の抽出操作・前処理に重要となる有機化学、(3) クロマトグラフィーなどによる分離などを含む分析科学、(4) データ変換・データマイニングによるインフォマティクスの操作に大別され、それぞれの操作に特有のノウハウが存在する(図 2)。それぞれのステッ

プにおける取り組みそれ自体が単独の研究として成立するほどであるが、本章では主に分析科 学とインフォマティクスに関する技術の現状と技術開発の取り組みを示す。

生命科学、有機化学、分析科学、インフォマティクスに基づく複合領域研究 organic acid alcohol lipid amino acid alkaloid alkaloid

図 2. メタボロミクス研究の流れ

2.1 ノンターゲット/ターゲットメタボロミクスの研究目的と用いる技術の違いについ τ

メタボロミクス研究は、生体サンプル内に"どの代謝物種"が"どのくらい"存在しているかという情報に基づいているという性質上、検出器によって観測された信号がどの代謝物によるものかを明らかにすること、すなわち代謝物同定が一つの大きなテーマとなってくる。メタボロミクスにおける代謝物同定(またはアノテーション)は、主にクロマトグラフィーによる溶出時間・検出化合物の吸収スペクトルや質量情報・核磁気共鳴(NMR)スペクトルなどの化合物情報の組み合わせにより行われる。しかしながら、現在、クロマトグラフィー質量分析(MS)などの種々の方法を用いて同定・定量が可能な代謝物種は 1,000 に満たないと言われている。これはこの世に存在するといわれている代謝物総数の 1 %未満[1]であるため、我々は限られた範囲での"網羅的な"解析を行っているというのが現状である。

そこで、なるべく多くの代謝物情報を失うことなく解析を行う手法としてノンターゲットメタボロミクスという手法が挙げられる。これは生体サンプルを分析して観測される成分全てを対象とする解析法で、網羅性という観点からは有用な方法論の一つである。しかしながら上述の通り、観測全ての成分に対して同定作業を行うことは至難の業である。そこで、メタボリックフィンガープリンティングという手法を駆使することにより、例えば測定サンプルの産地判別などの特定の目的に対して、化合物同定を省略しながらも達成することが可能となる。メタボリックフィンガープリンティングでは、検出器によって得られた信号パターンを直接、各種多変量解析(2.3. にて紹介する)などに供することにより、含有する代謝物パターンが類似しているサンプルをグルーピングしたり、クラス分けすることが可能となる。本手法は、検出器の信号強度のみに依存しているため、グループ分けに"どの代謝物"が寄与しているのかといった情報は度外視しているのが特徴である。また、代謝物同定が必ずしも必要ではないため、クロマトグラフィー質量分析に限らず、水素炎検出器(FID)や NMR、近赤外分光(NIR)などの成分情報を反映した出力結果が得られる全ての分析装置に適用可能である。こうした原因代謝物を特定しない手法は特に、産地判別の他、品質予測モデルの作成などにも威力を発揮す

る。分析機器に制約されないこうした研究は世界中で活発に行われており、当研究室においても、FT-NIR を用いた日本茶の品質予測[2]や、GC/FID を用いたチーズの官能評価予測モデルの作成[3]などにおいて非常に良好な結果を得ている(図 3)。

ターゲットメタボロミクスは上述のノンターゲットメタボロミクスと対をなす解析法であり、ある特定代謝経路の構成成分や生理活性物質、目的物質前駆体など目的成分に対象を絞った解析法であり、研究目的に沿って重要代謝物が前もって決まっている場合に用いられる。この解析法では、多成分中からいかに効率的且つ正確に目的成分を同定・定量するかが鍵となるため、対象化合物の保持時間情報と質量情報を同時に得ることが可能であるクロマトグラフィー質量分析法が多用される。特に、保持時間の再現性が高いガスクロマトグラフィーや質量分析装置の中でも定量性に優れる三連四重極型質量分析計などの組み合わせが多く用いられる。これらの解析では、事前に対象とする化合物の標準物質などを用いるなどして、一定分析条件下での保持時間や質量情報、及び化合物濃度とシグナルの関係を示す検量線などの in-house ライブラリを構築しておく必要がある。これらライブラリに格納された情報と実サンプルを分析して得られたピークと照会することによって、検体中に含まれる目的成分の同定及び定量を行う。最近では、標準化合物の保持時間(保持指標)情報や質量パターンなどが Web ページ上に公開されており、ライブラリデータベースの公共化を目指した取り組みも進められている[4][5]。

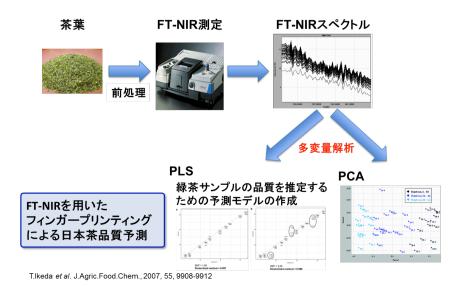


図 3. メタボリックフィンガープリンティングによる品質予測研究

2.2 質量分析法を用いたワイドターゲットメタボロミクス

ワイドターゲットメタボロミクスとは、ターゲット解析の定量性や再現性を確保しつつ、より目的成分を拡張し、同時に 100 以上の成分の代謝物定量を行うという手法であり、近年ではメタボロミクス研究の主流となりつつある方法論である。言うまでもなく、目的成分の拡張とデータ品質・スループットはトレードオフの関係にあるため、データ品質を保ったまま一方的に対象化合物を増やすことはできない。しかしながら、昨今の質量分析計の性能向上はめざましく、定量性に優れた三連四重極型質量分析計(QqQ/MS)を用いて数百成分を一斉検出したり、四重極と飛行時間型質量分析計を組み合わせた装置 (Q-TOF/MS) を用いて半ノンターゲット的な分析が可能となった。こうした装置性能の向上もワイドターゲットメタボロミクスを可能とした一因であることは間違いない。

この他にも質量分析機器のベンダーや研究者たちによって盛んに開発が進められている化合物の自動同定ソフトの貢献は非常に大きく、目的化合物拡張に伴う膨大な労力と時間消費の省

略に一役買っている。用いられているアルゴリズムは異なるが、実際にソフトウェア内で行わ れている作業は手動同定の際と同様、ライブラリに格納された化合物の保持時間情報や質量情 報と検出された化合物ピーク一つ一つを照合し、類似度を計算する。そして、ある一定以上の 類似度をもつピークを当該化合物として認識する、といったものである。この方法は計算機に よる演算の結果であり、手動同定につきものである解析者の主観が入り込む余地がないため、 客観的且つ再現性良く同定作業を行えるようになってきている。こうしたソフトウェアは解析 ソフトや質量分析機器のベンダーから有償で手に入れることができるが、研究機関に所属して いる研究者が開発したソフトウェアが無料で配布されている場合もある。当研究室では、津川 裕司氏(現理化学研究所所属)の開発した代謝物自動同定ソフトウェアである Aloutput を用い た GC/MS メタボロミクスを常用している。本ソフトウェアは理研 PRIME の Web ページ上で 無償配布されており、また約 500 代謝物の保持指標と質量スペクトル情報が格納されたライブ ラリも入手可能であるため、メタボロミクス技術導入の際に非常に役立つと考えられる [6]。 これらのソフトウェアを用いて自動同定された代謝物を用いて定量作業を行うことにより、前 もって目的代謝物を決めて解析を行う場合と比して広範囲の解析が可能となる。ソフトウェア を用いて自動同定を試みる手法は得られた全てのピークに対して機械的に解析を行うという工 程があるため、ノンターゲットメタボロミクスの一部であるとする場合もあるが、より広範囲 なメタボローム解析を可能とする技術の一つとして、節を分けてここに紹介した。

2.3 多変量解析を用いたデータマイニング

メタボロミクスにおけるデータ処理は、検出器より得られた信号強度の分布(クロマトグラムとして出力される)データを統計処理が可能な数値データに変換するところから始まる。クロマトグラフィーなどにより観測したピークを同定し、ピーク面積値を積分することにより、各代謝物のシグナル強度を含むデータマトリクスが生成される。このデータマトリクスを多変量解析に供することにより、データの視覚化が可能となる。ピーク同定作業の工程や現状については既に述べたので、ここではその後のデータマイニングについて言及する。

<データの視覚化の方法>

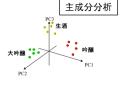
大きく3種類に分類、目的に合ったものを使用

探索的データ解析

<目的>サンプルのグルーピングや異常なサンプルを発見する例:メタボロームデータから数種類の日本酒の違いを識別できるか



主成分分析(PCA), 階層クラスター解析(HCA)など



回帰分析

<目的>ある変数を用いて別の変数を予測する

例:日本茶のランキング予測モデルにどんな化合物が寄与しているか



▶ 重回帰分析(MLR), 偏最小二乗回帰分析(PLS)など



判別分析

<目的>サンプルをカテゴリーごとに区別し、原因代謝物を探索する

例:微生物の分類にどの代謝物種が寄与しているのか



K最近傍分析(KNN), PLS判別分析(PLS-DA)など

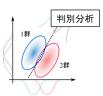


図4メタボロミクスに用いられるデータマイニング手法

ノンターゲット/ワイドターゲット解析にて得られたメタボロームデータは膨大な情報を含む ため、通常、多変量解析によって処理される。しかしながら、生体試料を分析して得られたデ ータは、分析を含む各ステップで系統誤差によるバイアスがかかっており、品質(信憑性)の 低いデータが混入している場合が多い。また、研究目的によっては、得られた代謝物データの ほとんどが意味のないものであり、真に意味のある代謝変動がこうしたノイズに埋もれて見づ らくなってしまっているという場合が往々にして起こり得る。こうした膨大なデータセットの 中から有益な情報のみを抽出してくる作業をデータマイニングと呼び、メタボロミクス研究の 中核を担う操作となる(図 4)。この作業は、未同定ピークにも適用可能であり、一旦未知ピ ークとしてアノテーションしておき、クラスター分離に寄与した未知ピークがあれば、集中的 に同定作業を行うという戦略も有効である。メタボロミクス研究に頻繁に用いられる多変量解 析によるデータマイニングの種類としては、主成分分析(PCA)、階層クラスター解析 (HCA)、重回帰分析(MLR)、偏最小二乗回帰分析(PLS)、判別分析(DA)、K 最近傍分 析(KNN)などがあり、得られたデータ構造や研究目的によってより良い解析法を選ぶ必要が ある。特によく用いられる多変量解析手法は、探索的データ解析としての PCA、HCA であ る。これらは膨大なデータ群の特性を調査し、類似サンプルのグルーピングやデータが含んで いる情報の性質を俯瞰することを目的としている。したがって、回帰分析や分類のモデル構築 を行う前段階として、扱っているデータセットの潜在的な可能性を確認することもできる。ま た、他にも MLR や PLS を用いた回帰分析などもメタボロミクス研究において重要となる。メ タボロミクスにおける回帰分析の究極的な目的は、代謝物データに基づき別の変数(病気の進 行度,お茶の品質など)を予測し、どの代謝物が原因代謝物であるかを明らかにすることであ る。さらに、サンプルをカテゴリーごとに区別し、その判別に寄与している代謝物を明らかに するのが DA や KNN などの判別分析である。最近では、サポートベクターマシンを用いた認 識精度の高い判別方法も開発されている。

本章では、メタボロミクスで用いられるデータマイニング手法を簡単に紹介した。それぞれの解析手法について独自のノウハウが存在するため、具体的なデータ変換操作や研究例については割愛するが、メタボロミクスにおける多変量解析は"データの視覚化"の手法にすぎないということをご留意いただきたい。これらの手法は、多変量のデータを人間の脳で扱うのが困難であるために、データ構造を低次元化・単純化し、視覚的にわかりやすい図に変換しているだけであり、ある課題に対して答えを与える性質のものではない。すなわち、多変量解析結果のみから結論を得ることはできず、関与が示唆された変数の別アプローチからのバリデーションを行うことが必要不可欠であると言える。

3. メタボロミクスの応用展開と展望

前章では、メタボロミクス技術の現状とその技術開発について分析・インフォマティクスを軸として紹介した。本章では、メタボロミクスによる細胞内代謝観測と、物質生産への応用例を中心として紹介する。二酸化炭素排出削減や持続可能型社会構築の観点から、微生物を用いたバイオプロセスで液体燃料や有用物質を生産する技術、及びその収率改善を目的とした種々の代謝改変戦略法の構築が世界中で進められている。微生物や植物を利用して効率的な物質生産を行うためには、物質代謝フローの最適化による目的プロダクトの生産能力の向上が重要となるが、そのためには代謝制御メカニズムを把握し的確に改変し制御することが必須となる。生体内に含まれる代謝産物を網羅的に解析するメタボロミクス研究は、目的物質生産と関連している重要代謝経路の抽出と律速となっている反応を探索するのに威力を発揮するため、バイオ物質生産の分野においても注目されている。

3.1 安定同位体希釈を用いた代謝物絶対定量

従来のメタボローム解析による表現型解析では、主に細胞内代謝物の相対濃度に基づき株間や培養条件間での差異を議論してきた。しかしながら、酵素反応は熱力学及び量論によって支配されており、物質生産における律速段階を正確に推定するには、細胞内の代謝物間の絶対量

比較の議論が重要となる。一般に、サンプルマトリクスによるイオン化抑制(サプレッション)や抽出時の誤差などを鑑みると、外部標準法による絶対濃度決定には限界があると言われており、同位体希釈法などを用いたより精度の高い手法が要求される。同位体希釈法は、分析対象化合物と同じ化学性質をもつ内標準物質を用いるため、理論的には、抽出操作から質量分析に及ぶほとんどの系統誤差を補正し得る。しかし、同位体希釈法において内部標準として用いられる安定同位体標識の標準物質は非常に高価である場合が多く、かつ商用利用可能なものに制限されるため、標的代謝物を拡張しているワイドターゲットメタボロミクスには不向きである。こうした問題を解決するため、安定同位体希釈に用いる内部標準物質を安定同位体で完全標識された細胞抽出液で代替するという手法が考案され[7]、特に微生物メタボロミクスの分野で広まりつつある。この方法では、内部標準として用いる細胞抽出液の濃度を事前に決定しておく必要があるが、検体と同じ細胞を用いることにより、検体で検出可能な代謝産物を網羅できるというメリットがある(図 5)。

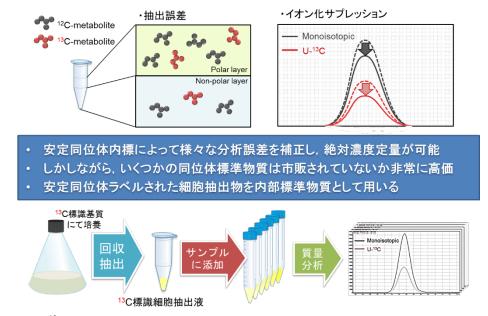


図 5. ¹³C-標識細胞抽出物を用いた安定同位体希釈法による絶対濃度測定

このようにして細胞内代謝物の絶対濃度を決定することにより、代謝産物間の量の比較が可能となり、さらに濃度という単位に変換しているため、古今東西の代謝物データとの比較が可能である。これは、研究機関や機器間に互換性のないシグナル強度のみのカタログと比べると非常に有益なデータ形式であるといえる。また、細胞内代謝物濃度から代謝中間体の関わる反応の標準エネルギーが計算でき、当該反応が発/吸エルゴン反応であるかが明らかとなるため、代謝経路上の律速反応の推定を行うというアプローチも存在する[8]。

3.2 安定同位体追跡に基づく動的代謝プロファイリング・代謝経路探索

代謝とは、本質的に動的な性格を持っており、常に水が流れている川のようなものであるととらえることができる。この流れの大きさを代謝流速と呼び、この流れの大きさ、向きを正確に観測することが、細胞の生理状態細胞を理解する一助となる。これは物質生産を論じる上でも同じことが言え、目的物質への代謝フローの最適化が一つの大きな課題となってくる。上述の通り、熱力学的なアプローチを用いることにより律速段階の推定が可能となるが、代謝物を静的なものとしてとらえる従来のメタボローム解析では動的な代謝の流れを捉えきれない場合がある。このような場合に対応するため、安定同位体追跡は、より直接的に代謝の大きさ(代

謝フラックス)を観測する手法として開発が進められてきた。特に代謝フラックス解析は、生物の生理状態の違いを代謝という側面から理解し、また、有用物質生産菌に適用することにより、生産性増大が期待できる実用化技術となっている。しかしながら、代謝フラックス解析はゲノム情報が未知である実用生物などには応用不可であるため、我々のグループは、数理的計算を必要としない、動的代謝プロファイリング法を代謝解析や物質生産研究に用いている。動的代謝プロファイリングとは、同位体標識された基質を細胞に取り込ませ、基質由来の各代謝物の同位体標識パターンの経時変動を、メタボローム解析技術により網羅的に解析するというものである。細胞内の定常状態が必須ではなく、また用いる標識基質の制約がないという点において比較的簡便な代謝動態解析法であると言える(図 6)。

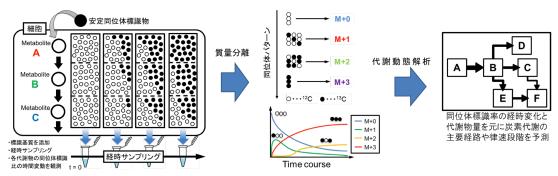


図 6. 動的代謝プロファイリングのワークフロー

この方法論は、基質やサンプリングの方法を工夫することにより、臓器のような培養細胞でない複雑な構造体を構成している細胞にも適用可能である。また、光合成チャンバー内で二酸化炭素を標識源として用いることにより植物体のまま動的代謝プロファイリングを行い、C3型光合成経路の律速反応を見出したという例もある [9]。 さらに、同位体標識基質から流入した炭素は全て標識されるため、物質生産における予期せぬ副生成物の発見や、目的生産フローに不必要な代謝経路の特定にも役立つことが期待される。近年、安定同位体追跡によって得られた独特のデータマトリクスを処理するためのソフトウェアや同位体パターンを認識する質量分析計の機能も開発されてきており、本手法の技術的なハードルは年々低くなってきている。したがって、今後は物質フローの最適化による物質生産研究だけでなく、様々な分野にまで代謝動態解析が応用・実用化されるということが予想される。

4. おわりに

本稿では、メタボロミクス技術の現状とその応用について概説した。近年、メタボロミクスの技術開発、及び本研究領域は全体として成熟してきたように思われるが、分析機器の発展や新たなソフトウェア性能の向上によって新たなブレークスルーが起きる余地は十分にあるように感じられる。また、分析技術の一般化や解析の自動化の流れ、そしてソフトウェア・公共データベースの充実はメタボロミクスの技術的ハードルを下げ、多くの研究者の参入を加速させるものと予想される。そして相対定量から絶対濃度定量への流れ、安定同位体追跡による代謝観測技術の発展は、これまで明らかにされなかった生命現象のより一層の理解に貢献することを期待したい。

謝辞

執筆の機会を与えていただきました日本薬学会構造活性相関部会に厚く御礼申し上げます。

参考文献

- [1] Fernie, A. R., Trethewey, R. N., Krotzky, A. J. and Willmitzer, L. Metabolite profiling: from diagnostics to systems biology, *Nature Reviews Molecular Cell Biology*, **5**, 763-769 (2004)
- [2] Ikeda, T., Kanaya, S., Yonetani, T., Kobayashi, A., Fukusaki, E. Prediction of Japanese green tea ranking by Fourier transform near-infrared reflectance spectroscopy, *Journal of agricultural and food chemistry*, **55**, 9908-9912 (2007)
- [3] Ochi, H., Bamba, T., Naito, H., Iwatsuki, K., Fukusaki, E. Metabolic fingerprinting of hard and semi-hard natural cheeses using gas chromatography with flame ionization detector for practical sensory prediction modeling, *Journal of Bioscience and Bioengineering*, **114**, 506-511 (2012).
- [4] http://www.massbank.jp/
- [5] https://metlin.scripps.edu/index.php
- [6] http://prime.psc.riken.jp/Metabolomics Software/Aloutput/index.html
- [7] Bennett, B. D., Yuan, J., Kimball, E. H., Rbinowitz, J. D. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach, *Nature protocols*, **3**, 1299-1311 (2008)
- [8] Nishino, S., Okahashi, N., Matsuda, F., Shimizu, H. Absolute quantitation of glycolytic intermediates reveals thermodynamic shifts in Saccharomyces cerevisiae strains lacking PFK1 or ZWF1 genes, *Journal of Bioscience and Bioengineering*, **120**, 280-286 (2015)
- [9] Hasunuma, T., Harada, K., Miyazawa, S., Kondo, A., Fukusaki, E., Miyake, C. Metabolic turnover analysis by a combination of in vivo ¹³C-labelling from ¹³CO₂ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C₃ photosynthetic pathway in Nicotiana tabacum leaves, *Journal of Experimental Botany*, **61**, 1041-1051 (2010)

植物におけるオミックス統合ネットワークシステム生物学

理化学研究所 福島敦史

1. はじめに

植物は、その進化の過程で、多種多様な低分子化合物を生合成する代謝経路を発達させてきた。生命維持に必要な一次代謝物に加えて、植物の持つ多様な二次代謝物群は、医薬品の原料としても重要である。近年の計測技術の革新は、植物の生長と発達メカニズムの背後にある分子実体に関して多くの情報をもたらしている [1]。特に、すべての遺伝情報(ゲノム塩基配列)の解読は、モデル植物のみならず作物種にも急速な広がりを見せている。高速かつ網羅的な新型配列シーケンサは、転写物や細胞内遺伝子の発現機構に関する情報(例.遺伝子発現データ)をも与える。質量分析計等の分析機器の発達により、転写物の翻訳後にできるタンパク質や酵素による代謝、細胞内分子間の相互作用に関する情報も比較的低コストで得られるようになった。

このような網羅的な分子情報を組み合わせ、その相互作用に焦点を当てて、植物の生長と発達の理解を目指す分野を植物システム生物学と呼ぶ。植物システム生物学の最終目的の一つは、遺伝型―表現型関係性の解明である。分析機器から得られた分子プロファイルデータから細胞内分子のふるまいを推定する。例えば、遺伝子発現データを用いた遺伝子機能推定は、重要な研究課題の一つである。

本稿では、転写物および代謝物プロファイリングの現状とそれらのデータ共有について概観 し、データ解析に用いられるネットワーク解析を紹介する。著者らの研究を通して、その背景、 ネットワークの構築、データ解釈について植物システム生物学の一端を解説する。

2. 高速かつ網羅的分子プロファイリングを支える分析機器の進展

「ゲノム」という用語は、遺伝子(gene)と全体を意味するオーム(ome)との組み合わせであり(gene + ome = genome)、遺伝情報の総体を意味する。ゲノムの研究分野自体はゲノミクス(genomics)と呼ばれる。同様に、転写物あるいは代謝物ならば、それぞれトランスクリプトーム(transcript + ome = transcriptome)とトランスクリプトミクス(transcriptomics)、メタボローム(metabolite + ome = metabolome)とメタボロミクス(metabolomics)となる。ここでは、植物科学においても発展の著しいトランスクリプトーム解析とメタボローム解析の現状を紹介する。

2.1 RNA シーケンシング (RNA-seq)

Illumina 社に代表される高速塩基配列シーケンサ [次世代シーケンサ (Next generation sequencer 略して NGS)]は、ゲノム配列解読のみならず、細胞内のゲノム DNA から転写された主要な RNA 分子群のプロファイリングにも利用される (RNA sequencing 略して RNA-seq)。 RNA-seq は、ゲノム情報の無い非モデル植物の遺伝子発現解析も可能なため、急速に普及した。従来、網羅的な遺伝子発現解析には、核酸のハイブリダイゼーションを原理とした DNA マイクロアレイが使用されてきた。これは、標的となる核酸配列があらかじめそのチップ上に搭載されていない限り、転写物レベルの検出は不可能であった。一方、RNA-seq では、新規転写物、small RNA、選択的スプライシング (alternative splicing)の検出が可能となった。明らかにRNA-seq は、植物の生長と発達を理解する上で欠くことのできない情報をもたらす技術の一つとなっている。著者の所属する研究グループでも、薬用植物に焦点を当てた RNA-seq による比較トランスクリプトーム研究を展開している。例えば、アカジソとアオジソとの間では、主に

色素アントシアニン生合成経路に関わる遺伝子群の発現に差が見いだされ、既知の情報と共に幅広い遺伝子発現様式に関する情報をもたらした [2]。

RNA-seq 解析は大きく分けて、ライブラリ作成、シーケンシング、データ解析からなる。NGS プラットフォームは常に改良がなされ、データの質と量共に発展し続けている。例えば、1 分子シーケンサは従来よりも長い配列を読み、転写物の全長を解読できる。また、組織別の発現解析や 1 細胞トランスクリプトーム解析も実現可能になってきた。機器の発達に加えて、RNA-seq ライブラリ作成手法の発展も目覚ましい [3]。例えば、Illumina 社の TruSeq RNA sample preparation kit によると、通常 3 日間で 24 ライブラリの作成が可能であった。これが、最近開発された Breath Adapter Directional sequencing(BrAD-seq)[4]によると、6 時間で 96 ライブラリの作成が可能となっている(それに伴ってコストも大幅に減っている)。シーケンサからは、数億~数十億もの短い配列を大量に取得できる(FASTQ 形式のファイルで数十 GB のサイズ)。このような大量データに対し、より高速かつ高精度な配列アセンブル法や発現量の定量方法の開発が急務である。デファクトスタンダードであるアルゴリズムを搭載し、ユーザフレンドリな GUI を備えたソフトウェアも整備されてきている。植物の持つ有用形質(多くの場合、量的形質)に資する遺伝子制御ネットワークの解明のために、トマトやトウモロコシといった様々な作物に対し、遺伝学と RNA-seq とを組み合わせたアプローチが盛んに行われている。

2.2 代謝物プロファイリング

メタボロームとは、生体内にある低分子代謝物の総体を指し、メタボロミクスは生物システ ム中の代謝物を網羅的に研究する分野をいう。しかし、一度の測定でサンプル中の代謝物すべ てを測定できるプロトコルも分析機器も無い。現状では、検出できる代謝物カバー率を上げる ために、複数機器を組み合わせて用いる [5]。植物界に存在する代謝物は 20~100 万種とも推定 されているが、現在の分析機器がカバーできているのは全体のごく一部である。これは、本質 的に代謝物の化学的多様性に起因する。現在、ガスクロマトグラフ(gas chromatography、GC) や液体クロマトグラフ(liquid chromatography, LC)と質量分析計(mass spectrometry, MS)とを 組み合わせた機器 (GC-MS と LC-MS) や核磁気共鳴がよく用いられる。一度に測定できる代 謝物数とデータ精度とは、一般的にトレードオフの関係にある。標的を絞った微量代謝物の検 出(絶対定量)を目的とした分析プロトコルは、単一もしくはごく少数の代謝物の存在量を信 頼性高く決める。より幅広い複数の化合物クラスの一斉検出を目的とするプロトコル(「代謝 物プロファイリング」=現状のメタボロミクス)では、検出できる代謝物数は増える一方で、 その絶対濃度情報が犠牲になる。現状で、GC-MS は揮発性成分のみならず、アミノ酸、糖、有 機酸、脂肪酸といった幅広い一次代謝物を同時に測定できる。LC-MS は、脂質に加えて、フラ ボノイド、グルコシノレート、フェニルプロパノイド、アルカロイド、ポリアミンといった主 要な植物二次代謝物群のプロファイリングに用いられる。両機器による代謝物プロファイリン グは再現性が高く、互いに相補的な代謝物カバー率を有するため、よく組み合わされて植物研 究に用いられている。

代謝物プロファイリングは、主に分析化学および分離技術を駆使した分析、データ取得とその解析からなる。この 15 年で、代謝物プロファイリングのために様々なデータ解析ツールやデータベースが開発されてきた [1,5,6]。近年では、KEGG [7]や KNApSAcK [8]といった化合物データベースや PlantCyc [9]などパスウェイデータベースに加えて、代謝物プロファイルに関するデータベースの開発も進んできた。Plant & Microbial Metabolomics Resource (PMR) [10]は、140 種を越えるモデル植物シロイヌナズナ変異体の代謝物プロファイルデータを提供している。PMR と相補的な情報として、我々の研究グループでも 50 種のシロイヌナズナ変異体の GC-MSによる代謝物プロファイルデータを公開した [11]。これらデータベースには、変異体やその遺伝子情報に加え、植物体の写真、代謝物情報、統計解析結果の視覚化など様々なデータとツールが含まれる。分析プロトコルやマススペクトル情報、生データファイルを提供している点でも、代謝物プロファイルデータの共有に向けた試みの一例となっている。

2.3 オミクスデータの共有

ゲノム研究のもたらした効果の一つは、データの共有と考えられる。歴史的には、ゲノム時代以前から研究コミュニティ内での塩基配列およびアミノ酸配列データの共有は進んでいた。加えて、NGS など新技術の出現とその大量データが、標準化と再利用性の確保とを目的としたオープンサイエンスの動きを促進した。例えば、マイクロアレイ実験に関する MIAMI (Minimum Information About a Microarray Experiment) は、かかる実験に関して最小限かつ基本的な情報を記述する枠組みであり、データ解釈がしやすく、得られた結果を独立に検証できるよう促している。これに準拠した形式で、NCBI の GEO (http://www.ncbi.nlm.nih.gov/geo/) はマイクロアレイや RNA-seq データなどの遺伝子発現データを収集・公開している。同様に、代謝物プロファイルデータに関しては、the Metabolomics Standard Initiative (MSI) による標準化に向けた努力がなされ、生データのリポジトリとして、MetaboLights [12]が利用可能である。このようなデータ共有の利点は何であろうか?当然、研究の再現や検証には役立つ。教育目的での利用も可能である。公開されたオミクスデータは、元々の研究プロジェクトの論文内容を支えるものであるから、品質は安定している。このようなデータセットから、さらなる知識発見が期待されている。実際に植物分野でも、このような公的に利用可能なデータセットを用いて、遺伝子機能を予測するデータベースの開発が盛んである [3, 13, 14]。

3. ネットワーク解析

ここでは遺伝子発現データを例にして、ネットワーク解析を解説する。遺伝子の相互作用は、ネットワーク(数学的にはグラフと呼ばれ、頂点と辺から構成される)として表現できる。このようなネットワークの概念を用いて、ゲノムスケールの細胞内分子間相互作用を定量化し、統合的に生物プロセスを理解していくアプローチをネットワーク解析と呼ぶ[1]。特に、公共利用可能なデータセットなどを用いて、遺伝子機能に関して検証可能な仮説を立てる、いわゆるデータ駆動型のアプローチに焦点を当てる。

3.1 典型的オミクスデータの解析

遺伝子発現データは、様々なデータ前処理や正規化法が施された後、発現データ行列として取り扱われる(図 1A)。このデータ行列は、遺伝子とサンプルからなる(通常、遺伝子数 p が数万のオーダであるのに対して、サンプル数 n は少数であり、いわゆる p >> n)。よくある比較は実験群(例.発変異体)と対照群(例.野生型)の 2 群間で、発現レベルの平均値に差がある遺伝子を統計的に調べることである。例えば、全遺伝子に対して、t 検定を繰り返し用いて、発現変動遺伝子(Differentially Expressed Genes, DEGs)を同定する(図 1B)。また、DEGs は遺伝子オントロジー(Gene Ontology, GO)タームエンリッチメント解析、Gene Set Enrichment Analysis(GSEA)あるいはその他のパスウェイ解析のような下流解析の出発点となる [6]。これらの手法は、あらかじめ特定した DEGs が、どういった機能を持つ遺伝子が多いのか(あるいは少ないのか)を調べるデータ解析法である。このような DEGs を同定する手法とその発展的な解析手法は、数多く開発されており、関連ツールも豊富である。ゆえに、データ解釈がこれら DEGs リストに過剰なほど依存する傾向にあることは否めない [15]。一般に、DEGs に関する GSEA やその他パスウェイ解析が有効となるには、既知の知見が使えるか否かが重要である。また、2 群間で発現量に差があったパスウェイ中の遺伝子のうち、どれが重要かについての情報が必ずしも得られるとは限らず、注意が必要であろう。

3.2 遺伝子共発現とは何か?

3.1 で述べた DEGs の同定とパスウェイ解析に加えて、(例えば、公共利用可能なデータを集めた場合に)複数サンプルにわたった 2 つの遺伝子の発現パターンを、類似尺度により特徴づけることができる(共発現,図 1C)。類似尺度の例の一つは、ピアソンの相関係数であり、遺伝子発現パターン間の関連性の一つを表す。数学的には、ピアソン相関係数は、変数間の直線関係のみに着目しており、(直線以外の)非線形関係の検出には向いていない。さらにデータの外れ値にも弱い。スピアマンの順位相関係数は、外れ値に影響を受けにくく、単調増加や単調

減少といった非線形関係は推定できる。このような尺度は、探索的データ解析では強い武器であり、線形・非線形関係のいずれも検出できる統計量の開発研究が行われている。

複数のサンプルにわたって、ある遺伝子とある遺伝子の発現パターン間の相関係数が高いとき、これらを「共発現している」とみなす。このとき、遺伝子群を頂点として表し、遺伝子間で共発現している場合(閾値を設ける)にはそれら頂点間を辺で結ぶ(共発現ネットワーク)(図 1D)。このような共発現計算を行い、共発現ネットワークを構築したとき、これらのネットワークはどのような特徴を持つだろうか。共発現ネットワークの特徴の一つは、自然界や人工物のネットワークにもよく見出される「スケールフリー(scale free)性」を示すことである。すなわち、少数のハブ遺伝子(頂点が持つ辺の数=次数が多い)が存在している一方(図 1E)、大多数の遺伝子はごくわずかの次数しか持たない。ハブ遺伝子は、致死性のタンパク質との相関や、非同義置換と同義置換の比による相対的な進化速度との関連が報告されている。シロイヌナズナの場合、ハブ遺伝子の進化速度は、非ハブ遺伝子に比べて相対的に遅い[16]。

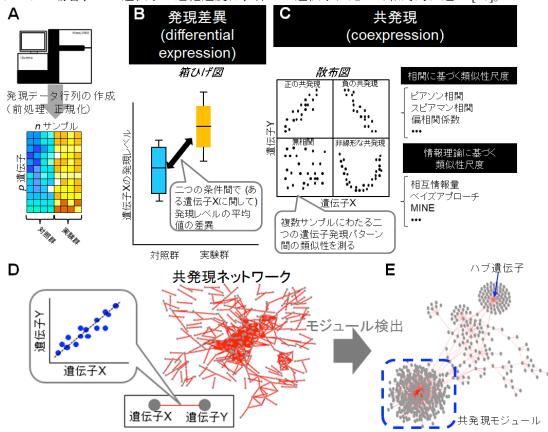


図 1. 典型的なオミクスデータの解析方法。マイクロアレイや RNA-seq で得られたデータは、遺伝子発現データ行列(典型的には、遺伝子数 p >>サンプル数 n)へと変換され、データ解析の出発点となる(A)。実験群(例. 変異体)と対照群(例. 野生型)の 2 群間で,ある遺伝子 X の発現レベルに差があるか否かは、t 検定などで調べられる(B)。図中の箱ひげ図は、各群における遺伝子発現データのばらつきを視覚化する。複数サンプルにわたった 2 つの遺伝子発現パターンは、類似尺度を用いて特徴づけられる(共発現,C)。散布図を用いることで、2 つの遺伝子発現パターンの関係性が視覚化される。ある遺伝子発現パターン群間の相関係数が高いとき、これら遺伝子群は共発現しているとみなす。このとき、遺伝子群を頂点として表し、共発現した遺伝子群を表す頂点間を辺で結んだ共発現ネットワークを構築できる(D)。共発現ネットワーク中で密につながりあった共発現モジュール同定のイメージ(E)。図中のハブ遺伝子とは、頂点が持つ辺の数(=次数)が多い遺伝子を指す。

3.3 共発現ネットワークモジュールの同定

共発現関係は、生物学的に意義のある遺伝子相互作用である場合と、非生物学的な変動に影響を受けた場合とがある。データのみから、これら変動の違いを数学的に区別することは一般的に難しい。前者の場合には、一例として転写因子による発現調節が考えられる。転写因子がある一団の下流遺伝子群のプロモータ領域に特異的に結合し、転写調節を行う場合には、これら下流遺伝子群同士は共発現しやすいという仮説が成り立つ。mRNAの分解、染色体上での遺伝子位置、エピジェネティクなヒストンアセチル化・メチル化、miRNAの標的同士も共発現関係を生じる重要な因子であろう。一方、後者の非生物学的な(意味のない!)共発現関係が生じる原因のいくつかは、RNAの品質やbatch effect に帰する。batch effect とは、サンプルを得る時期が異なった場合や別の研究室で行われた場合に、データ取得の際に発生する実験品質の違いがオミックスデータに生じることを指す(特に大規模な実験では、不可避な変動である)。通常、これらバイアスの推定と影響の緩和は、実験デザインに基づいた線形モデルによって行うことができ、重要視されるようになっている。実際、我々も50種のシロイヌナズナ変異体のメタボロームデータ解析の際には、このようなbatch effect を考慮した[11]。

共発現関係が生じる理由を踏まえて、共発現解析による遺伝子機能予測の考え方を紹介する。 共発現ネットワーク中には、よく共発現しており、互いに密につながりあった遺伝子群がよく 現れる(共発現モジュールや共発現クラスタと呼ばれる)(図 1E)。これら共発現モジュールに 含まれる遺伝子群の大半に同じ機能注釈(アノテーション)が付与される場合、残りの機能未 知遺伝子もその機能に関連する可能性がある。すなわち、発現パターンが似ている遺伝子群は、 その機能や役割が似ている可能性が高いという推定 ["guilt-by-association"(連座の誤謬)] に基 づいて、遺伝子機能予測が可能となる。すなわち、機能が既知である遺伝子情報をガイドとし て、未知機能遺伝子を探索できる。このようなアプローチは、特に二次代謝物の生合成経路 (例. フラボノイドやグルコシノレート) に関わる酵素遺伝子や、その制御に関わる転写因子 を同定する手法の一つとして利用されてきた [13,14]。

3.4 ディファレンシャルネットワーク

図 1B の発現差異(DEGs) および共発現解析(図 1C) の自然な拡張として、ディファレン シャル共発現(共発現差異、以降 differential coexpression, DC)なる解析手法を導入できる(図 2A)。すなわち、2 つの群間で各々測定された遺伝子群間の共発現パターンは、異なる場合があ る。広義には、異なるネットワーク間で変化したつながり(あるいは相互作用)を、システム レベルで明らかにする手法として、ディファレンシャルネットワークとも呼ばれる[17]。対照 群と実験群の 2 群間で共発現関係に違いが無い場合は、条件によらず常に共発現している遺伝 子群とみなせる (ハウスキーピング共発現ネットワーク)。一方、2 群間で、有意に異なる共発 現値を示す遺伝子群を同定することで、DC ネットワークを構築できる。このようなディファ レンシャルネットワークを同定するアルゴリズムは複数提案されており [18]、著者も Fisher's Z-test に基づくアルゴリズムを実装した DiffCorr パッケージを開発した [19]。DC ネットワーク は、様々な細胞内システムの調節機構の変化が反映されていると考えられる。例えば、DC の 原因の一つは、転写因子のプロモータ領域の多型である(図 2B)。従来の DEGs 同定に比べる と、この手法は 1 遺伝子の変化ではなく、遺伝子間の相互作用における変化に着目する。これ は細胞内である特定の分子(群)の生理活性を制御するエフェクタや、制御因子自身がある変 異によってその作用効果が変化する場合であっても、自身の遺伝子発現には変化が無い場合に 有効である。著者らは、公共利用可能なトマトのマイクロアレイデータセットを中心として、 共発現モジュール同定と DC 解析を行った [20]。例えば、葉と果実サンプル間で大規模な DC ネットワークを比較し、代謝経路のキーステップである遺伝子機能予測に役立てた。このよう に、従来の DEGs や共発現解析では未探索であった鍵酵素遺伝子や転写調節因子候補を、大量 データから抽出できることが示唆される。

一般的に、同定した DC ネットワークは巨大である。遺伝子機能解析のためには、何らかの優先順位づけがさらに必要である。近年、発現差異と DC とを組み合わせ、転写因子等の制御因子情報を活用したディファレンシャルネットワーク解析法が発展している[15]。この手法は、

制御因子をコードしている遺伝子が、2 群間で有意な DEGs(大きく変化した遺伝子群)と共発現差異するか否かを調べる。これら情報は重みづけされ、ある制御因子のネットワーク変化への影響を反映した統計量に変換される。このような統計量を用いて、すべての制御因子候補について順位づけが行われる。植物では前例がないが、動物ではこのようなアプローチがすでに行われている。一例として、筋肉質なピエモンテ牛と和牛由来の発達ステージにわたるマイクロアレイデータから Regulatory Impact Factor (RIF) なる統計量を算出したとき、筋肉質筋肉肥大の原因遺伝子ミオスタチン(Myostatin; MSTN、骨格筋分化抑制遺伝子)が正しく推定された[15]。AtGenExpress プロジェクト [21]により、モデル植物シロイヌナズナを乾燥ストレス処理し、その遺伝子発現応答を調べたマイクロアレイデータが利用可能である。これを用いた著者の予備的調査によると、これまでよく知られている乾燥ストレスに関わる因子群のいくつかは比較的高い RIF 値を持っていることがわかった(未発表)。本手法が植物データに有効であるか否かは、今後の詳細な検討が必要である。オミクスデータ解析において、分子存在量の増減のみならず、ネットワーク全体からその役割や協調関係性に焦点を当てたゲノムワイドな評価が重要であると考えられる。

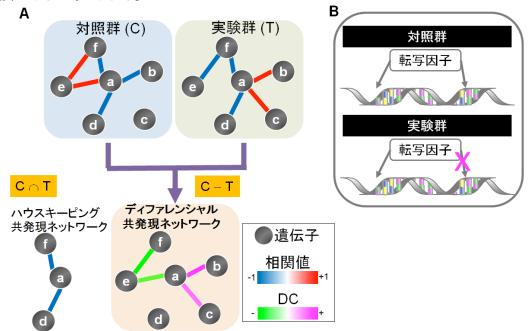


図 2. ディファレンシャル共発現 (DC) ネットワークの概念図。対照群と実験群の 2 群間で共発現関係に違いが無い場合は、条件によらず常に共発現している遺伝子群が存在 することを意味する (ハウスキーピング共発現ネットワーク) (A)。一方、2 群間で統計 的に有意に異なる共発現値を示す遺伝子群からなる DC ネットワークを構築できる。この DC ネットワークは、様々な細胞内システムにおける調節機構の変化が反映されている。 例えば、転写因子のプロモータ領域の多型 (×印) は、DC の原因の一つと考えられる (B)。

3.5 ネットワーク解析の勘所

いうまでもなくネットワーク解析は万能ではない。しかるべき実験計画に基づく高品質なデータの取得が大前提となる。遺伝子機能解析につながるデータ解釈の際、はまり易い落とし穴も多い。遺伝子発現データを用いたネットワーク解析を例にして、以下に注意すべき点を述べる。

(1) 発現差異(DEGs)に過度に依存した解析と解釈には注意が必要。

- (2) 相関係数値だけに固執しない。例えば、ピアソン相関係数値は、外れ値データから極端に 影響を受ける。散布図を確認することは重要である。
- (3) 相関関係と因果関係は必ずしも同じではない。偶然同じ発現パターンを見ているだけかもしれない(と疑ってかかる方が良い)。
- (4) 制御因子の発現レベル自体は必ずしも高いわけでない、転写レベルではなくむしろその活性が転写後レベルで調節されていることも考慮する。この場合、DEGs 同定や共発現パターンのみでは、標的たる制御因子にたどり着くことは困難。
- (5) 共発現だけでなく、ディファレンシャル共発現の情報をもっと活用すべきである。ただし、 DEGs 同定に比べて、多くのサンプル数が必要なので、資金繰りが難しい?

4. おわりに

今後、高速かつ網羅的な分析技術がさらに発展し、それに伴って、さらなるバイオインフォマティクスと生物統計の手法開発が必要となる。DEGs を同定する手法やその発展手法の数に比べて、ディファレンシャルネットワークを調査できるソフトウェアは決して多くない。特に、GUIによるユーザフレンドリなソフトウェアの研究開発の発展が望まれる。

ネットワーク解析によるオミクスデータ解釈は、当然ながらすべての実験的・技術的なデータ産出とデータ加工の影響を受け、遺伝子機能予測の精度に影響する。遺伝子機能が実験的に特徴づけられている割合は、シロイヌナズナですら全遺伝子の 20%程度である。ゆえに、精度評価に使える gold-standard と呼べる遺伝子セットの利用は難しい。また、植物種に共通した(ロバストな) 共発現パターンとは何なのか?共発現解析に適したデータセットとは何か?進化的に保存されたディファレンシャル共発現関係が存在するのか?ネットワーク解析から導かれる制御因子候補の優先順位づけは、どの程度当たるのか?といった疑問は尽きない。

公的データベースへのデータ公開と共有は、公開データからの知識の再発見および検証可能な仮説構築を促す効果を発揮することは想像に難くない。これまでの研究財産である公的利用可能なオミクスデータ群が宝の山であるとするならば、ネットワーク解析がこのような用途に有力な手がかりを与えてくれると信じている。

謝辞

本稿執筆にあたり、多くの共同研究者の方々との研究成果を参考にしました。この場を借りて厚く御礼申し上げます。筑波大学の草野 都先生には丁寧にご査読いただき、貴重なご意見とご助言とをいただきました。心より御礼申し上げます。本稿で紹介した著者らの研究の一部は、科学研究補助金・若手研究 B (26850024) の支援を受けたものです。また本稿執筆の機会を与えて下さった日本薬学会構造活性相関部会の先生方に深謝申し上げます。

参考文献

- [1] Fukushima A, Kusano M: A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. *J Exp Bot*, **65**, 5619-5630 (2014).
- [2] Fukushima A, Nakamura M, Suzuki H, Saito K, Yamazaki M: High-Throughput Sequencing and De Novo Assembly of Red and Green Forms of the Perilla frutescens var. crispa Transcriptome. *PLoS One*, **10**, e0129154 (2015).
- [3] Ichihashi Y, Fukushima A: Frontiers of Transcriptomics in Plant Science. *BSJ-Review (in Japanese)*, **7C**, 110-123 (2016).
- [4] Townsley BT, Covington MF, Ichihashi Y, Zumstein K, Sinha NR: BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Front Plant Sci*, **6**, 366 (2015).
- [5] Fukushima A, Kusano M: Recent progress in the development of metabolome databases for plant systems biology. *Front Plant Sci*, **4**, 73 (2013).
- [6] Fukushima A, Kanaya S, Nishida K: Integrated network analysis and effective tools in plant systems

- biology. Front Plant Sci, 5, 598 (2014).
- [7] Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, **44**, D457-462 (2016).
- [8] Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin M, Darusman LK *et al*: KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol*, **53**, e1 (2012).
- [9] Chae L, Kim T, Nilo-Poyanco R, Rhee SY: Genomic signatures of specialized metabolism in plants. *Science*, **344**, 510-513 (2014).
- [10] Hur M, Campbell AA, Almeida-de-Macedo M, Li L, Ransom N, Jose A, Crispin M, Nikolau BJ, Wurtele ES: A global approach to analysis and interpretation of metabolic data for plant natural product discovery. *Nat Prod Rep*, **30**, 565-583 (2013).
- [11] Fukushima A, Kusano M, Mejia RF, Iwasa M, Kobayashi M, Hayashi N, Watanabe-Takahashi A, Narisawa T, Tohge T, Hur M *et al*: Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development of a Metabolite Profiling Database for Knockout Mutants in Arabidopsis. *Plant Physiol*, **165**, 948-961 (2014).
- [12] Kale NS, Haug K, Conesa P, Jayseelan K, Moreno P, Rocca-Serra P, Nainala VC, Spicer RA, Williams M, Li X et al: MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Curr Protoc Bioinformatics*, **53**, 14 13 11-18 (2016).
- [13] Kusano M, Fukushima A: Current challenges and future potential of tomato breeding using omics approaches. *Breed Sci*, **63**, 31-41 (2013).
- [14] Yonekura-Sakakibara K, Fukushima A, Saito K: Transcriptome data modeling for targeted plant metabolic engineering. *Curr Opin Biotechnol*, **24**, 285-290 (2013).
- [15] Hudson NJ, Dalrymple BP, Reverter A: Beyond differential expression: the quest for causal mutations and effector molecules. *BMC Genomics*, **13**, 356 (2012).
- [16] Chae L, Lee I, Shin J, Rhee SY: Towards understanding how molecular networks evolve in plants. *Curr Opin Plant Biol*, **15**, 177-184 (2012).
- [17] Ideker T, Krogan NJ: Differential network biology. Mol Syst Biol, 8, 565 (2012).
- [18] Kayano M, Shiga M, Mamitsuka H: Detecting Differentially Coexpressed Genes from Labeled Expression Data: A Brief Review. *IEEE/ACM Trans Comput Biol Bioinform*, **11**, 154-167 (2014).
- [19] Fukushima A: DiffCorr: an R package to analyze and visualize differential correlations in biological networks. *Gene*, **518**, 209-214 (2013).
- [20] Fukushima A, Nishizawa T, Hayakumo M, Hikosaka S, Saito K, Goto E, Kusano M: Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches. *Plant Physiol*, **158**, 1487-1502 (2012).
- [21] Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K: The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*, **50**, 347-363 (2007).

オミクス統合解析を用いた 植物二次代謝機能ゲノミクスについて

マックスプランク分子植物生理研究所 峠 隆之

1. はじめに

植物の代謝物は大きく一次代謝物と二次代謝物に分類される。前者は生体を維持するのに必 須の化合物群と定義され人類は植物からそれらを栄養素として摂取しているのに対して、後者 は生物種固有のものが多く、そのうちの多くが植物の防御機構に関わる生物活性物質群として 知られ薬用成分などとして用いられている。私たちはその植物二次代謝物の中でも、フラボノ イド(ポリフェノール類の一種)と総称される赤~紫色素であるアントシアニジンや淡黄~白 色のフラボノールやイソフラボノン等に代表される化合物群(図 1)の生合成経路に着目して 研究を行っている。歴史的にフラボノイド生合成経路に関わる遺伝子群は、主に自然変異体を 用いたフォーワードジェネティクス(順遺伝学)的手法により発見されてきた。特に遺伝子欠 損が視覚的表現型として現れるアントシアニジン生合成経路遺伝子は、トウモロコシや花の自 然色素変異体の解析などから多数同定された(図 1 (i) 参照)。またいくつかの酵素遺伝子は、 代謝物の化学構造から予想できる酵素反応に着目した活性試験を指標に、タンパク質抽出液か ら標的酵素を分離およびアミノ酸配列を解読することで同定されてきた(図 1 (ii) 参照)。さ らに、こうして特徴付けされた生合成経路に関わる遺伝子の塩基およびタンパク質のアミノ酸 配列情報をもとに、同植物内の相同遺伝子(ホモログ遺伝子)や他種における相同性遺伝子 (オルソログ遺伝子)、および共通したドメイン配列を有する遺伝子群に着目することで、同 一・類似した機能もしくは新規機能遺伝子の発見が行われてきた。さらに 2000 年、種子植物 として初めてシロイヌナズナのゲノム解読が完了して植物科学におけるポストゲノム科学が開 始して以来、T-DNA 挿入変異体などの人為突然変異体を用いて、種皮色素(アントシアニジン 類の重合体)の欠損を指標にした大規模な変異体の作製と視覚的スクリーニングが行われ、数 多くのフラボノイド生合成関連遺伝子が機能同定された(図 1(iii)参照)。しかし、上記の研 究により解明されてきた生合成遺伝子群は、遺伝子の欠損が生合成経路上の大部分の代謝物欠 損、すなわち視覚的表現型に関わる早期段階生合成酵素遺伝子群(EBG: early biosynthetic genes) (図 1 参照)、転写因子群および輸送系関連遺伝子群が主であった。しかし配糖体化やアシル 体化など、フラボノイドの化学安定性および構造多様性に関わる修飾酵素遺伝子(LBG: late biosynthetic genes) の解明は、酵素反応の複雑性と多様性の理由から関連遺伝子の発見が非常に 難しかった。

一方でポストゲノム科学の開始以来、ゲノム配列情報から遺伝子総体(オーム)を基盤として研究を行ういわゆる消去法型の研究手法が可能となったため、このようなオーム科学を前提とした研究の考え方すなわちオミクス科学が、ゲノム、mRNA(トランスクリプトミクス)、タンパク質(プロテオミクス)および代謝物(メタボロミクス)それぞれのレベルで利用可能となった。さらに、それぞれのオミクス科学を統合解析という形で合わせ見ることで、より詳細かつ包括的な理解が可能であるという考えのもとに、様々な組み合わせのデータ統合解析の試みがなされた。しかし植物二次代謝生合成経路に関しては、その二次代謝物の化学構造の多様性・複雑性から、ゲノム情報からの生合成経路の全体像の推定は困難であることに加え、生合成経路の細胞内局在の複雑性などの理由により、メタボロミクスデータを他のオミクスデータとの統合解析において消去法的に用いるのは困難であった。

そこで本稿では、遺伝子 - 代謝ネットワークを用いたオミクス科学の統合による機能ゲノミクスのいくつかの研究例を紹介する。遺伝子 - 代謝ネットワークとは、一部では生合成経路自体のことを示す際に用いられる場合もあるようだが、一般的には転写制御因子の発現量とその

制御下の代謝物の蓄積レベル、さらには遺伝子の発現量と代謝物の蓄積量の相関関係を議論する際に用いられる単語である。本稿では機能ゲノミクス(新規遺伝子機能の発見・同定)を行うために、転写制御因子の過剰発現体、生合成経路の器官特異性の活用、自然変異体の解析および植物種間比較などを用いて、メタボロミクスを基盤とした統合解析を試みた研究例を紹介する。多くの場合は研究標的となる該当遺伝子はゲノム上に一つしかなく、もしデータや仮説になにか間違いがあった場合、後の確認実験が全て無駄になるため、データや候補遺伝子の絞り込みは非常に注意深く行う必要がある。各章では私たちが直面した問題や後に判明した問題など、現実的に起こりうる問題点に触れつつ、どのように新規遺伝子の特定に至ったかを含めて紹介する。

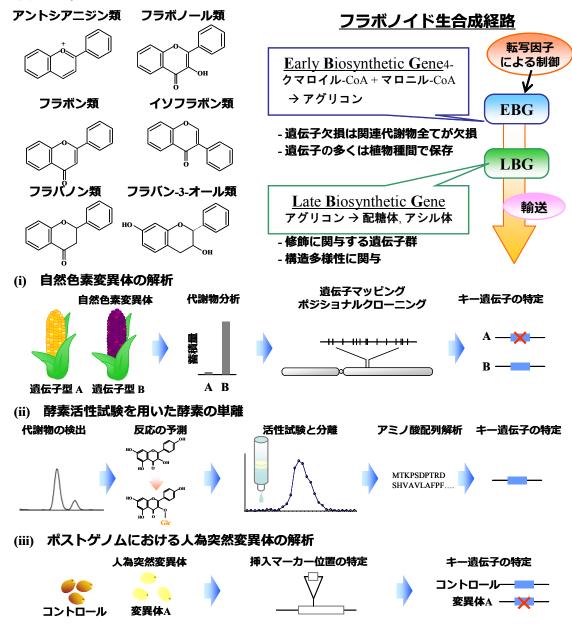


図 1 種子植物におけるフラボノイド生合成経路と関連遺伝子の機能同定の歴史について

左上部:主要フラボノイド類。右上部:フラボノイド生合成経路遺伝子群の概略。

2. メタボロミクス的手法を用いた標的生合成経路の全体像の把握

植物二次代謝機能ゲノミクスの研究例を紹介する前に、一つの大きな問題点について触れておきたい。メタボロミクスは当初、代謝物全てを同時に解析するメタボロームを起点として行う研究の手法・考え方であった。しかし、塩基配列やアミノ酸配列とは違い、化学物性の違う代謝物を全て網羅する分析は技術的に非常に難しいことから、現在主に用いられているメタボロミクスという言葉のニュアンスは少し違う。すなわち現状では、メタボロミクスとは『可能な限り網羅的な代謝物分析』を行うことを指し、その範囲内で他のオミクスデータと統合する際に用いられる言葉になっている。ここでの『可能な限り網羅的』とは、少し曖昧な印象を与えるかもしれないが、標的生合成経路の全体像を捉えるために様々な動的データ(ストレス応答性や器官特異性など)を含めて解析を行い、生合成経路の全体像を出来る限り把握することを前提とする。そのため、目的にもよるが、統合解析に用いる他のオミクスデータは、代謝物解析に用いた同様もしくは類似した検体を用いたほうがより精度の高い解析結果が得られる。一方で、分析で見落としている代謝物に関する生合成経路については、オミクス統合解析のから情報を得ることはできない。そのため、研究目的に沿った実験計画が非常に重要である。

3. 転写因子過剰発現体を用いたトランスクリプトームデータとの一対一統合解析

トランスクリプトームデータとメタボロミクスデータとの統合解析は比較的歴史があるといえるが、その当時に利用可能であった情報の精度や範囲、および得られるデータの質などで、研究背景は大きく異なり、その時々でそれぞれに応じた試みや問題解決がなされていた。植物科学分野におけるトランクリプトーム解析といえば、初期のExpressed Sequence Tag (EST) やcDNA ライブラリー、ディファレンシャルディスプレイ法やマクロアレイ、また近年ではマイクロアレイや次世代シーケンシングなどが挙げられる。難しい点としては、質の良いデータを得るための技術革新を待つ間に、質の良くないデータから発見された新規遺伝子が論文として報告されれば、そこで研究として競争相手に打ち負けてしまうということである。その反面、質・精度のあまり良くないデータや、目的の明確ではない実験において統合解析を行うのは、非常に心折れる作業であり、どのような比較解析が明確にかつ特異的に標的生合成経路を標的として解析できるか実験計画を練ることが重要な点であった。しかし、近年の次世代シーケンス技術の発展は、こういった諸問題の多くを排除することができ、はるかに多くの生物種に応用可能であり汎用性も高いため、今後しばらくは主要な手法になると考えられる。

本章ではトランスクリプトームデータとの統合解析例として、生合成制御転写因子を用いた方法を紹介する。一般的に、代謝生合成経路は転写制御もしくは転写後の制御による調整や基質/産物による正/負のフィードバック制御および分解系などによってその産生・蓄積が調整されている。しかし、植物二次代謝経路の多くは転写因子制御により一方向(正)もしくは二方向(正と負)に比較的単純に制御されているため、分解系が働く生育条件下などの特別な場合を除いて、基本的に関連遺伝子の発現レベルが代謝物蓄積レベルによく相関することが知られている。現在までに、フラボノイド生合成を特異的に制御する MYB や bHLH などの転写因子が数多くの植物種で特定されていたので、これら転写制御因子の過剰発現体を用いて標的生合成経路を特異的に高発現させて統合解析に用いた例についていくつか紹介する。

3.1 モデル植物シロイヌナズナのアントシアニン転写因子 (PAP1) 過剰発現体を用いた 解析

この研究は、機能獲得型変異体であるアクティベーションタグコレクションから単離されたアントシニン過剰蓄積変異体として単離された pap1-D (Production of Anthocyanin Pigment 1-Dominant, AtMYB75) [1]を用いた統合解析の研究例である。研究を始めた当時は、この変異体について詳細な解析はあまり行われておらず、アントシアニンが高蓄積していることといくつかのアントシアニン生合成経路遺伝子が高発現していることが報告されていただけであった。

そこで、寒天培地で生育させた pap1-D 変異体のロゼット葉と根について、高速液体クロマト グラフィー質量分析計 (LC-MS) を用いた代謝物分析を行い、合計 11 種類のアントシアニン 誘導体の蓄積を認めた。当時唯一化学構造が同定されていたシロイヌナズナのアントシアニン 誘導体(A11:cyanidin-3-O-[2"-O-(2"'-O-(sinapoyl)-xylosyl)-6"-O-(p-O-(glucosyl)-p-coumaroyl)glucoside]-5-O-[6""-O-(malonyl)-glucoside]) の化学構造情報を元に MS/MS 分析を用いて詳細に 解析し、シロイヌナズナアントシアニン生合成経路の全体像を把握した。次にマイクロアレイ によるトランクリプトーム解析を行い、pap1-D変異体で高発現している遺伝子群を見出した。 これら変異体で高発現していた遺伝子群を生合成経路の全体像上で統合した結果、全ての既知 アントシアニン EBG 遺伝子群が認められたことに加え、いくつかの LBG 候補遺伝子が得られ た。さらに他の植物種で機能同定されているフラボノイド糖転移酵素遺伝子との配列相同性解 析を行い、これら候補遺伝子について、それぞれの生合成経路の酵素反応に対する機能推定を 行った[2]。得られた候補遺伝子と推定した遺伝子機能は、候補遺伝子の変異体を用いた代謝物 分析で in-vivo での機能の同定、および組換えタンパク質を用いた酵素活性試験で2つのアント シアニン LBG 遺伝子を新規機能遺伝子として同定した[2]。さらに後に続けられた研究で、4 つ の遺伝子についても推定された機能で同定された[3,4]。この統合解析手法は(図2参照)、標的 代謝物が特異的に高蓄積した変異体を用いてトランスクリプトームデータとの統合解析を行う ことで、機能ゲノミクスを行った研究である。後に知ったことであるが、別の研究グループも 同様の変異体を土壌条件で生育させてマイクロアレイ解析を行っていた。彼らの結果を見せて もらったが、高発現していた遺伝子群のリストが大きく違った。これはアントシアニンの高蓄 積による光合成阻害などの二次的な影響が強く影響していたことがわかった。このように、特 異的変化を用いて統合解析を行う場合は、生育条件などにも注意が必要である。

(iii) 転写因子過剰発現体を用いた解析

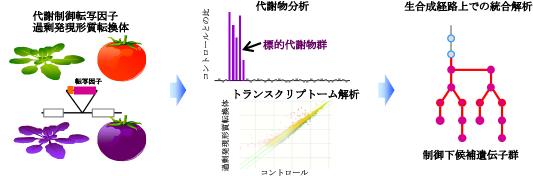


図 2 転写因子過剰発現体を用いた解析によるアントシアニン生合成経路関連遺 伝子の特定

3.2 モデル作物トマトのアントシアニン過剰発現体(Del/Ros1)を用いた解析

モデル植物シロイヌナズナを用いて成功した研究手法を作物モデルに応用する試みは非常に重要であることから、前章で述べた転写因子の過剰発現体を用いた統合解析を、果実作物モデルであるトマトへ応用した。トマト種においてアントシアニン生合成はあまり研究されておらず、化学構造も明らかとなっていないことがわかった。そこで、形質転換紫トマト(キンギョソウから単離されたアントシアニン制御転写因子である Del (MYB) と Ros I (bHLH) の過剰発現形質転換トマト) [5]の果実からアントシアニンを単離し、構造決定を行った。構造決定されたトマトアントシアニンは、ナスニンとペタニンとして知られているナス科特異的アントシアニンであった。これらの結果を、形質転換紫トマトを用いて次世代 RNA シークエンスによるトランスクリプトームデータと統合解析した結果、トマトアントシニン生合成経路に関わる全ての LBG 酵素遺伝子を特定することができた。さらに、その中でも機能新規性が高いナス

科特異的フラボノイドアシル基転移酵素遺伝子(SIFdATI)の機能同定に成功した[6]。この研究は、オミクス統合解析を用いた機能ゲノミクスがトマトのような作物モデル植物にも応用可能であること示した、重要な研究例であると考えている。

4. 公共遺伝子発現データとの統合

植物科学においては、最近 15 年ほどのトランスクリプトーム解析の発展とデータ蓄積とと もに、国際コンソーシアムによって組織的にデータベース上でトランスクリプトームデータの 公開・共有化が進んだ。また、GeneVestigator (https://genevestigator.com/gv/) や eFP browser (http://bar.utoronto.ca/) などの器官別およびストレス誘導性遺伝子発現解析を、ウェブ上で誰 もが簡単に解析できるようになった。これらの解析ツールは実験デザインの参考に用いること ができるのみならず、その公開されているデータを in silico データ解析として論文に直接使用 することが一般的になった現在の状況に貢献したといえる。さらに、公開された大量データを 用いて相関係数を算出し、遺伝子-遺伝子間の共発現解析ツールが数多く公開された。酵母など の研究分野から派生した共発現データ解析法は公開当初は遺伝子間の相関係数リストであった が、現在ではデータベース上でネットワークとして表示されるまでに至った。メタボロミクス との統合解析に用いる遺伝子共発現ネットワークは、基本的に標的生合成経路の既知遺伝子群 を用いて骨組みを構築し、転写レベルで発現様式が似ている遺伝子群を新規候補遺伝子として 見出す手法である。遺伝子共発現ネットワークデータベースの機能ゲノミクスへの貢献は多大 であり、細胞壁、フラボノイド、グルコシノレート、リグニン、タンパク質・アミノ酸分解経 路、自食経路など様々な生合成経路において機能ゲノミクスにおける研究成果が報告されてい る[7]。注意すべき点としては、共発現ネットワーク解析は相関係数算出に使用したデータセッ トの内容や組み合わせに大きく依存するため、既知遺伝子群をポジィティブコントロールとし て構築されたネットワークを評価する必要がある。転写レベルで発現挙動が類似しているスト レス誘導性や器官特異性などを標的にする場合、データ種類と組み合わせにより、非常に有効 な手段となる。最近ではシロイヌナズナのみならず、様々な植物において ATTED-II (http://atted.jp/) などで利用可能である。

4.1 遺伝子共発現ネットワークとの統合解析によるフラボノール生合成経路の解明

フラボノール(淡い黄色のフラボノイドの一種)はアントシアニンとは異なり、その蓄積お よび欠損が視覚的表現型に直結しないなどの理由から、シロイヌナズナにおいても構造や関連 LBG 酵素遺伝子など、解明されていない点が多かった。一方で、ロゼット葉や根において観察 されるフラボノール配糖体蓄積パターンはあまり複雑ではなかったため、初期の ATTED(当 時はシスエレメントの推定を主な目的としたデータベース)を用いた共発現ネットワークと代 謝物解析データと統合解析することで、2 つのフラボノール特異的酵素遺伝子の機能を新規に 同定した[8.9]。しかしその後、それぞれのシロイヌナズナの器官を分析した際に、フラボノー ル配糖体の蓄積パターンに多様な器官特異性が観察されたため、この複雑性に着目し、その器 官特異的フラボノール産生に関わる遺伝子群の機能同定を目的として、研究を始めた。しかし その器官特異性と多様性は予想以上に複雑であり、LC-MS で検出された新規化合物を含むと予 想される 40 種類ほどのフラボノールの構造推定は、MS/MS 解析や標品による検出ピークの同 定では困難であった。この問題を打開するために生物学的視点による新しい手法、すなわち生 合成遺伝子変異体を用いた構造推定法を見出した。これは酵素遺伝子の変異体はその下流代謝 物を全て欠損することを利用した方法である。遺伝子機能同定が既になされているフラボノイ ド酵素遺伝子の 10 以上の変異体について、器官別の代謝物分析を詳細に行い各検出ピークの 構造推定を行うことで、器官特異的新規フラボノールを含むほぼ全てのフラボノール類の構造 推定を行うことができた。その蓄積パターンを元に、シロイヌナズナにおけるフラボノール生 合成経路の全容を構築し、その代謝物蓄積パターンの器官特異性とともに関連遺伝子群の発現 様式の推定をおこなった。

その後、共発現データベースとして開発が進んでいた ATTED-II を用いて遺伝子共発現ネットワークを構築し、フラボノール生合成関連遺伝子と発現相関の高い機能未知候補遺伝子群を絞り込んだ。得られた候補遺伝子群をフラボノール生合成経路の全体像と統合することでほぼすべての修飾酵素反応の候補遺伝子、および反応基質の供給に関わる一次代謝関連遺伝子および輸送に関わる候補遺伝子など多数推定することができた(図 3 参照)。得られた候補遺伝子は、相補組み換え形質転換体および組み換え酵素活性試験などの実験により、いくつか機能同定を行った[10]。遺伝子共発現ネットワークは、用いたトランスクリプトームデータセット内での転写レベルでの相関解析であるため、用いたデータセットに大きく依存するが、特異的な転写制御下にある遺伝子群を絞り込む手法としては非常に強力である。

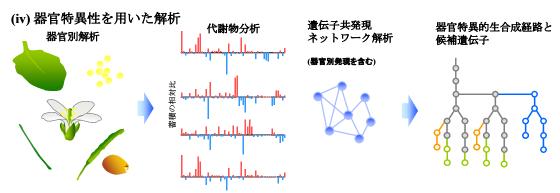


図 3 器官特異性と共発現ネットワーク解析の統合解析による器官特異的酵素候補遺伝子の特定

5. ゲノミクスとの統合解析:自然変異の解析と種間比較解析

これまでの章では、主にトランスクリプトミクスとメタボロミクスの統合解析を機能ゲノミクスに用いた例を紹介した。本章では、最近盛んに研究されている近縁種間比較におけるゲノミクスを基盤とした解析とメタボロミクスとの統合解析例について簡単に紹介したい。ここでの近縁種とは、交配が可能な植物種間を指し、自然界でそれぞれの生息地域に固定していたアクセッションと呼ばれる自然変異体コレクションなども含まれる。基本的に近縁種間解析ではそれぞれのゲノム配列の類似性が高いことから、SNPs などの遺伝子多型のマッピングやプロファイリングを用いて、交配ライン作成や目的に応じた相関解析に用いられる。さらに、次世代シークエンシングが配列解析の主力となった現在、次々と完了していく植物ゲノム配列の解読により、新機能遺伝子の探索手法は大きく変わってきていると実感している[11]。そこで本章では、将来の種間比較解析に向けた解析法についても簡単に紹介する。

5.1 代謝物量的形質座位解析 (mOTL)

QTL (量的形質座位)に用いることが出来る交配ラインは、その作製法と交配様式の違いにより IL (Introgression Line)、NIL (Near-Isogenic Line)、RIL (Recombinant Inbred Line)、BIL (Backcross Inbred Line)などいくつかに分類される。QTL 解析は、基本的に 2 種類の近縁種を交配して作製されたラインを用いて自家受粉もしくは片側の近縁種との交配を繰り返し、ゲノム SNP マーカー解析と組み合わせて単離されたゲノムワイドなホモ接合体をラインとして用いた解析法である(図 4 (v) 参照)。QTL 解析自体の歴史は長く、主にバイオマスや作物の収量に着目した解析が主であったが、近年私たちは植物二次代謝産物に着目した mQTL 解析をシロイヌナズナ[12,13]やトマト[14]などで行った。基本的に代謝物 QTL 解析により得られた遺伝子群は、近縁種間での交配による機能付加に用いることのできる遺伝子群であり、代謝物産生に関わる育種におけるマーカー遺伝子となりうるという点は非常に重要であるといえる。

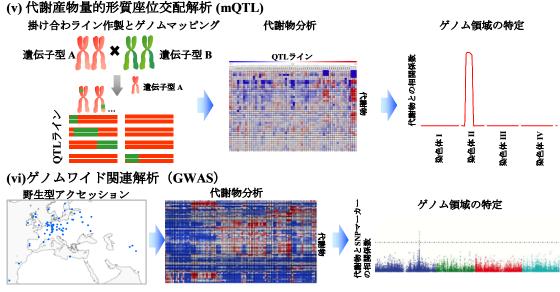


図 4 代謝物量的形質座位解析(mQTL)とアクセッションゲノムワイド関連解析(GWAS)

5.2 ゲノムワイド関連解析 (GWAS)

シロイヌナズナ野生型アクセッションの SNP プロファイルを用いた GWAS 解析について紹介する。近年シロイヌナズナでは、300 種類以上の野生型アクセッションの種子がバイオリソースとして利用可能である。これらのアクセッションは 20 万以上の SNP プロファイルのデータが既にデータベースに登録されているため、研究者の代謝物データを用いてウェブサイトで簡単に SNP マーカーとの相関解析を行い、候補遺伝子もしくは領域の絞込みが簡単に行える(図 4 (vi) 参照)。近年、私たちはフラボノールを指標とした GWAS 解析をシロイヌナズナ[12]で報告した。また 2016 年に 1135 種類のシロイヌナズナ野生型アクセッションのゲノム配列 が完了・公開されたため、遺伝子配列を活用した機能ゲノミクスに向けた新しい手法が可能であると考えている。

5.3 ゲノム配列を用いた種間比較解析

次世代シークエンシングによる配列解析により、現在までに 100 種類以上の植物種のゲノム配列が完了・公開されている。一方で、これらモデル植物種の多くは、トランスクリプトームデータの蓄積が少なく、メタボロミクスとの統合解析に用いるには不十分であるといえる。ゲノム配列とメタボロミクスの直接的な統合解析は非常に困難であるが、我々はゲノムデータを効果的に機能ゲノミクスに応用する将来に向けた試みを行っている。実際には、代謝物分析と文献検索から生合成経路の骨組みを構築し、近縁種との代謝物比較を行い、生合成経路を重ね書きして全体像を広げていく。さらに Plaza (http://bioinformatics.psb.ugent.be/plaza/) やPhytozome (https://phytozome.jgi.doe.gov/pz/portal.htm/) のようなゲノムワイドオルソログ遺伝子解析[11,15]を行い、遺伝子複製領域などを特定して植物種間の詳細なシンテニー領域解析を行う。この解析から得られた構造多様性に関わる遺伝子領域もしくは遺伝子群に関して、遺伝子発現解析や遺伝子相同性解析の結果と代謝物解析データと組み合わせることで、その機能ゲノミクスに向けた新しい戦略が可能であると考えている。

(vii) 種間比較解析を用いた二次代謝機能ゲノミクス

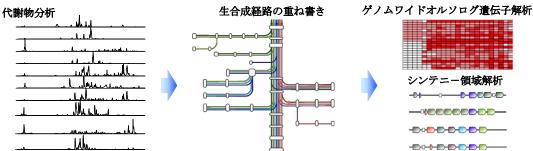


図5ゲノム配列を利用した種間比較解析による二次代謝機能ゲノミクスの手順

謝辞

執筆に際し機会を与えていただきました日本薬学会構造活性相関部会、ならびに貴重なコメントを頂きました編集員の先生方に感謝申し上げます。本稿で紹介した研究は千葉大学の斉藤和季教授、マックスプランク植物分子生理学研究所の Alisdair Fernie 博士と共著者およびラボテクニシャンの方々と実施されたものです。ここに深く感謝させていだきます。また、執筆の際にご教授くださいました渡邉むつみ博士に深く感謝させていだきます。

参考文献

- [1] Borevitz, J. O., Xia, Y. J., Blount, J., Dixon, R. A., Lamb, C., Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*, **12**, 2383-2393 (2000).
- [2] Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D. B., Kitayama, M., Noji, M., Yamazaki, M., Saito, K., Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J.*, **42**, 218-35 (2005)
- [3] Luo, J. Nishiyama, Y. Fuell, C. Taguchi, G. Elliott, K. Hill, L. Tanaka, Y. Kitayama, M. Yamazaki, M. Bailey, P. Parr, A. Michael, A. J., Saito, K., Martin, C., Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana. *Plant J.*, 50, 678-695 (2007)
- [4] Yonekura-Sakakibara, K., Fukushima, A., Nakabayashi, R., Hanada, K., Matsuda, F., Sugawara, S., Inoue, E., Kuromori, T., Ito, T., Shinozaki, K., Wangwattana, B., Yamazaki, M., Saito, K., Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana. *Plant J.*, **69**, 154-167 (2012)
- [5] Butelli, E., Titta, L., Giorgio, M., Mock, H. P., Matros, A., Peterek, S., Schijlen, E. G. W. M., Hall, R., D. Bovy, A. G., Luo, J., Martin, C., Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol.*, **26**, 1301-1308 (2008)
- [6] Tohge, T., Zhang, Y., Peterek, S., Matros, A., Rallapalli, G., Tandrón, Y., Butelli, E., Kallam, K., Hertkorn, N., Mock, H., Martin, C., Fernie, A., Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant J.*, **83**, 686-704 (2015)
- [7] Tohge, T., Scossa, F., Fernie, A. R., Integrative Approaches to Enhance Understanding of Plant Metabolic Pathway Structure and Regulation. *Plant Physiol.*, **169**, 1499-1511 (2015)
- [8] Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., Saito, K., Phytochemical genomics in Arabidopsis thaliana: A case study for functional identification of flavonoid biosynthesis genes. *Pure Appl. Chem.*, **79**, 811-823 (2007)

- [9] Yonekura-Sakakibara, K., Tohge, T., Niida, R., Saito, K., Identification of a flavonol 7-Ornhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J Biol Chem.*, **282**, 14932-41 (2007)
- [10] Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., Saito, K., Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell*, 20, 2160-2176 (2008)
- [11] Tohge, T., de Souza, L. P., Fernie, A. R., Genome-enabled plant metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci.*, **966**, 7-20 (2014)
- [12] Ishihara, H., Tohge, T., Viehöver, P., Fernie, A., Weisshaar, B., Stracke, R., Natural variation in flavonol accumulation in Arabidopsis is determined by the flavonol glucosyltransferase BGLU6. J Exp Bot., 67, 1505-1517 (2016)
- [13] Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., Fernie, A. R., Characterization of a recently evolved flavonol phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat Commun.*, 7, 12399 (2016)
- [14] Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., Kleessen, S., Giavalisco, P., Pleban, T., Mueller-Roeber, B., Zamir, D., Nikoloski, Z., Fernie, A. R., Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell*, 27, 485-512 (2015)
- [15] Tohge, T., Watanabe, M., Hoefgen, R., Fernie, A. R., The evolution of phenylpropanoid metabolism in the green lineage. *Crit Rev Biochem Mol Biol*, **48**, 123-52 (2013)

構造活性フォーラム 2016 開催報告

実行委員長 篠原 康雄 徳島大学先端酵素学研究所

構造活性フォーラム 2016 は 2016 年 6 月 24 日 (金)、淡路夢舞台国際会議場 (兵庫県淡路市) において、「分子標的薬の創生とゲノム創薬の目のつけどころ」という主題で開催されました。 最近、分子標的薬やゲノム創薬で創製される医薬品の有用性が高く評価されるようになって いることを受け、実用化に繋がるような創薬研究がどのようにして進められているのか、企業 やアカデミアの先生方にご紹介いただくことを今回のフォーラムの目的にしました。

ご講演いただいたのは以下の先生方です。

講演 1「キナーゼ創薬における構造活性相関の活用: キナーゼ選択性プロファイリングを用いたターゲットホッピング」 澤 匡明 (カルナバイオサイエンス株式会社)

講演 2「ゲノム創薬の実例 (バイオロジストの視点から): ALK 阻害剤アレクチニブ創製/開発を例にして」 青木 裕子 (中外製薬株式会社)

講演3「相互作用阻害によるがん抑制因子の再活性化を利用した 難治性乳がんに対する新規治療薬の開発」 片桐豊雅(徳島大学・先端酵素学研究所)

講演 4「カイコの感染モデルを用いた新規抗生物質の発見」 関水 和久 (帝京大学・医真菌研究センター)

カルナバイオサイエンスの澤先生からは、キナーゼ創薬の重要性について丁寧なイントロダクションを頂いたうえで、リード化合物の創出から、キナーゼ選択性の向上の取り組みについて、ご講演頂きました。また、中外製薬の青木先生からは ALK 阻害剤アレクチニブの創製、開発がどのように進められたかのご紹介を頂きました。とりわけ、個々の患者の分子や遺伝情報に基づいた「個別化医療」の重要性について具体例をご紹介いただけたのはとても魅力的でした。徳島大学の片桐先生からは、難治性乳がんに対するアカデミアからの取り組みについてご紹介いただきました。次回のフォーラムでトピックスとして取り上げられる予定のタンパク質間相互作用を標的として、巧みな創薬研究を進めておられる様子をご講演いただきました。また、最後の演者は、他の先生方の分野とは少し毛色が違ったのですが、カイコという、とてもユニークなモデル生物を用いて創薬研究を進めてこられた帝京大学の関水先生にライソシンEの発見についてお話を頂きました。

講演会の最後に、中外製薬の大田雅照先生がパネルディスカッションを担当下さいました。 個々の先生方の注目したところ、ならびに創薬研究を成功させるために鍵になった技術等をま とめてご紹介下さり、これらについて出席者と演者の先生方と意見交換を頂きました。

今回私どもがフォーラムを担当させていただくことになったのは、豊橋技術科学大学の高橋 由雅先生と大阪大学の高木達也先生から「2016年の構造活性フォーラムを担当頂けないだろう か」とご相談を受けたことに始まります。当初、私自身が全くの門外漢であることもあって躊 躇したのですが、自分の恩師である寺田 弘先生が構造活性相関研究に携わっておられる様子を

SAR News No.31 (Oct. 2016)

傍で拝見していたこともあって、「自分たちでお手伝いできるなら」とお引き受けすることになりました。実際には私自身は全くの「お飾り実行委員長」で、HP 開設、会告配布、会場の準備、運営やら、企業(15 社!)からの醵金集め、更には出席者のリクルート活動まで、私どもの研究室の山本武範講師が一手に引き受けて、うまく運営してくださいました。山本武範講師に感謝するとともに、本フォーラムがますます充実したものになっていくことを祈念しつつ、開催報告とさせていただきます。



写真はフォーラム 2016 の様子。67 名の出席者がそれぞれの演者の「目のつけどころ」に耳を傾けました。

<会告>

第44回構造活性相関シンポジウム:第31回農薬デザイン研究会との共同開催

主催: 日本薬学会構造活性相関部会,日本農薬学会農薬デザイン研究会

会期: 2016年11月16日(水)-2016年11月17日(木)

会場: 京都大学医学部創立百年記念施設 芝蘭会館

(〒606-8315 京都市左京区吉田近衛町 京都大学医学部構内

http://www.med.kyoto-u.ac.jp/shiran/)

日程:

11月16日(水)特別講演,招待講演,一般講演,ポスター発表,一般講演 合同懇親会(芝蘭会館 山内ホール)

11月17日(木)招待講演,一般講演

特別講演: 11月16日(水)

Isabelle Billas 博士 (IGBMC, France)

A Structural View of Allosteric Control of Transcription by Steroid Nuclear Receptors

招待講演:

- 1. 奥野恭史先生(京都大学大学院医学研究科)「スパコン・ビッグデータ時代の創薬」
- 2. 石川智久先生(個別化医療研究所)

「薬物トランスポーターの新しい構造活性相関解析法の開発と臨床・創薬への応用」

- 3. 福西快文先生(産業技術総合研究所) 「分子設計ソフトウェア myPresto による広範囲なツール提供: −精密分子設計の前 段階−」
- 4. 谷野圭持先生(北海道大学大学院)

「シストセンチュウふ化促進物質の全合成」

- 5. 森 直紀先生 (東京大学大学院・農学生命研究科) 「植物病害虫に関わる生物活性天然物の合成」
- 6. 池口満徳先生(横浜市立大学・生命医科学研究科) 「分子動力学シミュレーションと X 線溶液散乱法の連携によるタンパク質動的構造 解析」
- 7. 寺田 透先生(東京大学大学院農学生命科学研究科) 「分子シミュレーションで探るタンパク質-リガンド間相互作用」

Special Session: 11月16日(水)藤田稔夫先生の米寿お祝い

Thierry Langer (University of Vienna, Austria)

Computer-aided Molecular DesPign: From QSAR to Pharmacophore-based Ligand Profiling

ランチョンセミナー:

11月16日(水)アフィニティサイエンス+コンステラ11月17日(木)オープンアイジャパン

発表登録・参加申込など:

発表申込締切日:締め切りました 講演要旨締切日:締め切りました 参加申込締切日:10月14日(金)

シンポジウムホームページ http://www.qsarj.org/44sympo/index.html

参加登録費: [一般]予約 8,000 円 (当日 10,000 円)、[学生]予約 2,000 円 (当日 3,000 円)

懇親会参加費:[一般]予約 7,000 円、[学生]予約 3,000 円

問い合わせ先:

〒606-8502 京都市左京区北白川追分町

京都大学大学院農学研究科応用生命科学専攻

「第44回構造活性相関シンポジウム・第31回農薬デザイン研究会実行委員会」

中川好秋

e-mail: naka@kais.kyoto-u.ac.jp

URL: http://www.qsarj.org/44sympo/index.html

<会告>

3rd International Symposium for Medicinal Sciences

主催: 日本薬学会

会期: 2017年3月26日(日) 時間 未定

日本薬学会第 137 年会(仙台)の会期中

年会の会期は3月24日(金)~27日(月)

会場: 仙台国際センター会議棟

(〒980-0856 仙台市青葉区青葉山無番地、Tel. 022-265-2211)

主題: 新しい創薬ツール

招待講演

The Automation of Early Stage Drug Design Using a Robot Scientist

Professor Ross D. King

(Machine Intelligence, School of Computer Science at the University of Manchester)

King 博士とその共同研究者達は、創薬研究に応用可能な「人工知能ロボット化学者 Eve」を開発し、ハイスループットスクリーニングと、その結果を人工知能によって解析と検証を繰り返すことにより、新規抗マラリア薬のリード化合物の開発に成功した。この手法は、通常の新薬探索より、迅速で経済的にも優位な新薬開発の手法であることから、市場性の観点から製薬会社の対象とならなかった希少疾患の治療薬の開発に道を開く可能性を有している。

High resolution cryo-EM of protein complexes

Dr. Sriram Subramaniam

(Center for Cancer Research, National Cancer Institute, NIH)

Subramaniam 博士は電子顕微鏡法を用いて、タンパク質や受容体の構造を原子レベルの構造(2.2A の解像度)解析に成功した。この方法は、クライオ電子顕微鏡法(Cryo-EM)と呼ばれ、結晶化を必要とせず酵素と阻害剤等の低分子複合体の懸濁液を液体窒素で凍結させた後、電子線を当て画像撮影をし、4 万個の二次元投射イメージを撮影する。この画像をスパコンで三次元構造に再構成し、X 線結晶構造解析をしのぐ解像度でタンパク質の構造を「見る」ことに成功した。

Invited Poster Presentation

本年も、製薬企業、ベンチャー企業、アカデミアからポスター発表の依頼を計画いたしております。招待講演者の方々にも Poster Presentaion の討論に参加していただく様にお願いする予定です。特に本年は、上記招待講演の関連のテーマで発表していただける方にお願いする予定です。発表を希望される方は、下記まで連絡お願いいたします。

連絡先: 〒150-0002 渋谷区渋谷 2-12-15 公益社団法人 日本薬学会 3rd International Symposium for Medicinal Sciences 担当 横山祐作 E-mail: yokoyama@pharm.toho-u.ac.jp

部会役員人事

会計幹事が交代いたしました。新役員は以下の通りです。 会計幹事 前田美紀(農業・食品産業技術総合研究機構)

構造活性相関部会の沿革と趣旨

1970年代の前半、医農薬を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と分子設計に対する新しい方法論が世界的に台頭してきた。このような情勢に呼応するとともに、研究者の交流と情報交換、研究発表と方法論の普及の場を提供することを目的に設立されたのが本部会の前身の構造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980年からは年1回の「構造活性相関シンポジウム」が関係諸学会の共催の下で定期的に開催されるようになった。

1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。構造活性相関懇話会は1995年にその名称を同研究会に改め、シンポジウム開催の実務担当グループとしての役割を果すこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の構造活性相関に関する学術・研究の振興と推進に向けて活動している。現在それぞれ年1回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR News を年2回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。

本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。 (http://bukai.pharm.or.jp/bukai kozo/index.html)

編集後記

日本薬学会構造活性相関部会誌 SAR News 第31号をお届けいたします。

大阪大学大学院工学研究科 傳寶 雄大先生・福崎英一郎先生には、メタボロミクスの実験レベルの実践的な話題から新しい研究の方向性をお示しいただきました。理化学研究所 福島敦史先生には RNA レベルでの遺伝子を対象としたネットワーク研究をご紹介いただきました。マックスプランク分子植物生理研究所 峠 隆之先生には、代謝物、遺伝子、生合成経路を全体としてまとめて行く御研究を紹介いただきました。代謝物の同定から生合成遺伝子の制御あるいは未知遺伝子の機能同定などを様々な技術と手法論で紐付けて行く時代の変化に驚きました。同時に、我々の身近な技術の延長上にあることも理解しました。本誌第 29 号でご紹介いただいたような相互作用ネットワークの考え方も近い将来には組み込まれて行くように感じました。

第30号でも掲載致しました「構造活性相関シンポジウム(農薬研究会と共催)」と「薬学会での国際シンポジウム」についても情報が更新されておりますのでお目通しいただければ幸いです。

(編集委員会)

SAR News No.31 平成 28 年 10 月 1 日 発行:日本薬学会 構造活性相関部会長 高木 達也

SAR News 編集委員会

(委員長)飯島洋、小田晃司、河合健太郎、清田泰臣、田上宇乃、幸瞳

*本誌の全ての記事、図表等の無断複写・転載を禁じます。