

多変量解析および単変量解析を用いたNMRメタボロームデータの試料群間比較

関連製品：核磁気共鳴装置(NMR)

近年、NMRによって代謝産物の総体(メタボローム)を混合物のまま解析するNMR代謝プロファイリング(NMR metabolic profiling; NMR-MP)がポストゲノム研究や品質管理などで盛んに用いられています。NMR-MPIは、ノンターゲットメタボロミクスの一手法です。ノンターゲットの解析では、解析の対象が広範にわたるため、NMRスペクトルから直接データを解釈するのは困難です。多変量解析などを適用することで、多変量データから有用な知見を発掘できます(データマイニング)。

データマイニングでは、実験系および目的に合わせてデータの要約、分類、モデリング、回帰(予測)などの適切な多変量解析を適用することが重要です。ここでは、異なる試料群間において特長的な代謝産物(=マーカー分子)を探索したり、代謝産物をつかって試料間の判別モデルを作成することを目的に、試料の群情報(クラス分け)を目的変数として利用する多変量解析である(O)PLS-DAや単変量解析をNMRメタボロームデータに適用した例を紹介します。

試料と方法

異なる2つの試料群のモデル試料として、光環境および暗環境において生育したブロッコリスプラウトを用いました。ブロッコリの種子を1wt%寒天に播種し、播種から5日後まで、いずれも暗環境で生育し、播種から5-10日の期間は、光環境または暗環境で生育しました。播種から10日後に成育したブロッコリスプラウトを採取し、Bligh-Dyer法に準ずる方法で、代謝産物を抽出し、極性画分を遠心エバポレータで乾固しました。乾固した極性代謝産物を、重水で調製したpH = 7.0の100 mM リン酸カリウム緩衝液で再溶解させ、NMR試料としました。¹H-NMR計測には、シングルパルス法を用いました。それぞれの試料群につき8試料を生育し、NMR試料を作成しました。

それぞれのスペクトルの0.5-9.0 ppmの範囲を、等幅の積分範囲(0.04 ppm幅)で積分しました(バケット積分)。結果として194変数 * 16試料からなる行列(多変量データ)が得られました。OPLS-DAおよび単変量解析は、R言語におけるbioconductorライブラリ (<https://www.bioconductor.org/>) に記載されている"ropls"パッケージ [1] および"muma"パッケージ [2] (<https://CRAN.R-project.org/package=muma>) を用いて計算しました。

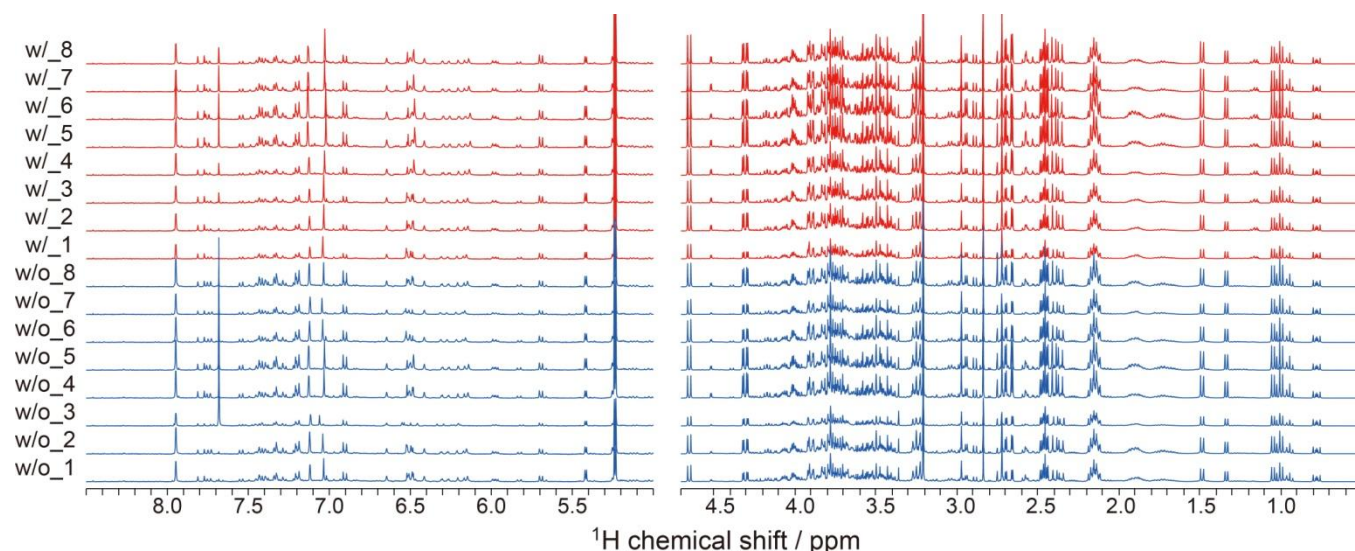


Figure 1. 二群のブロッコリスプラウトから抽出した極性代謝産物¹H-NMRスペクトル。5.0-8.5 ppmの領域は、Y軸を8倍拡大した。NMR分光計はJNM-ECZ400Sを、検出器はROYALプローブを用い、ASC30オートサンプルチェンジャーを用いて連続測定をおこなった。

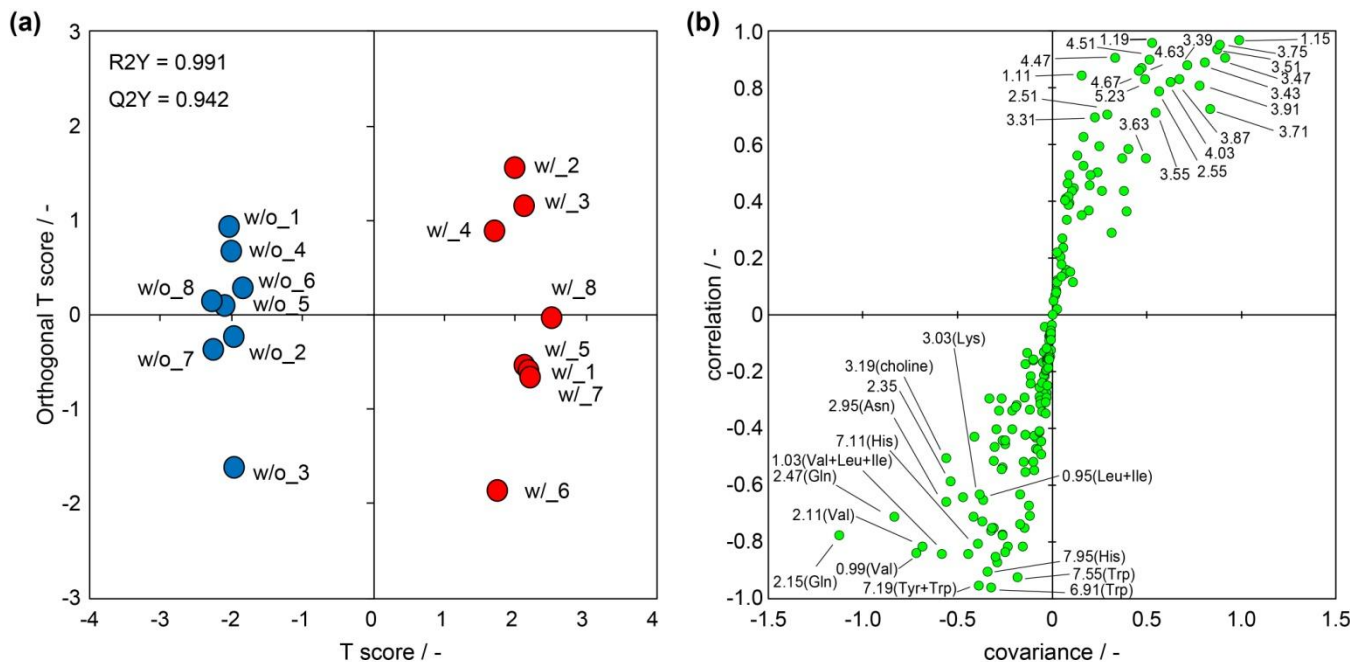


Figure 2. 二群のブロッコリスプラウトから抽出した極性代謝産物¹H-NMRスペクトルのOPLS-DA。(a) スコアプロット。(b) S-plot。スペクトルの総和でノーマライズし、中央化およびパレートスケーリングしたデータをもちいた。

判別分析の適用

NMR-MPでは、ほとんどの場合において、試料数に対して変量数の方が多くなります。そのような場合、線形判別分析を適用することは不適切です。このような場合は、部分最小二乗 (Partial Least Squares) モデルをつかった判別分析であるPLS-DAなどを用います [3]。線形判別分析では、群間の変動が最大になるように各変数の線形結合として判別関数を求めます。これは、NM170013 [4] で紹介したように、主成分分析と同様の手続きになりますが、主成分分析では、与えられた多変量から分散が最大となるように各変数の線形結合として主成分を作成します。しかしながら、データの分散を最大化する主成分が群間の変動を反映するとは限りません。PLS-DAでは、データと群情報に相当する目的変数の間の共分散を最大化するように、線形結合係数を選択します。したがって、得られる判別関数は群間の変動を最大化することができます。

本ノートでは、PLS-DAよりデータの解釈が容易なOPLS-DA [5] を適用しました (Figure 2)。各試料は変数の総和でノーマライズし、各変数は中心化し、スケーリングにはパレートスケーリングを用いました。

OPLS-DAのスコアプロットからは光環境と暗環境においてスコアが明確に分離していることが読み取れます (Figure 2(a))。作成されたモデルの信頼性の指標である説明変数による目的変数の説明度 (R²Y) および予測度 (Q²Y) は、それぞれ0.991および0.942であり良好な判別モデルが得られました。

S-plotは、スコアと変数の間の共分散 (x軸) および相関係数 (y軸) の散布図です (Figure 2(b))。したがって、第一象限には、光環境において特長的な代謝産物、一方、第三象限には、暗環境において特長的な代謝産物がプロットされます。第一象限は、グルコースやスクロースなどの糖が認められ、一方、第三象限には、グルタミン、アスパラギン、バリンなどのアミノ酸が認められました。これは、光環境での、光合成によって糖が豊富に供給される独立栄養的な代謝、および、暗環境での、炭素飢餓に起因するタンパク質のリサイクルによるアミノ酸の供給を示唆します [6]。実際には、NMR-MPからは因果関係は分からない (暗環境によるアミノ酸の増大がタンパク質のリサイクルに由来するかは分からない) ですが仮説を得ることができます (データ駆動型アプローチ)。因果関係を明らかにするには別に検証実験を構築する必要があります。NMR-MPなどのデータ駆動型アプローチを用いることで、従来の仮説駆動型アプローチでは、考え得なかった結論を得られる可能性を秘めています。

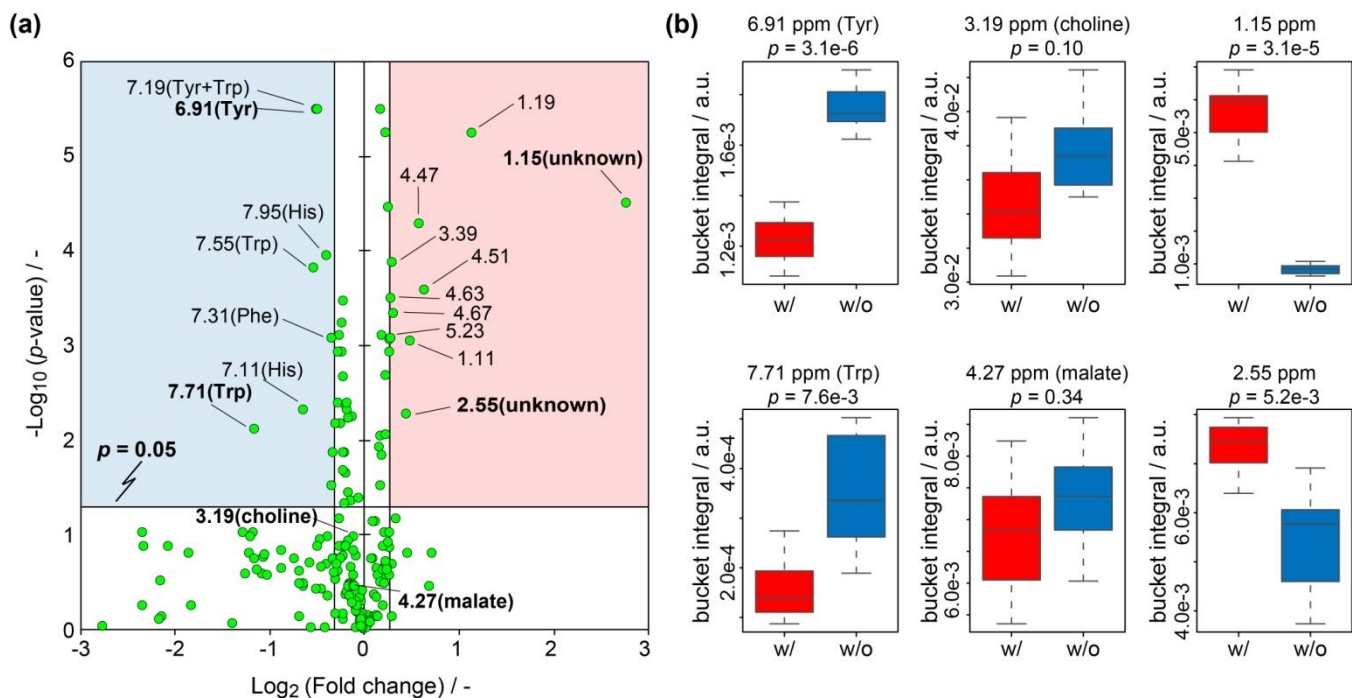


Figure 3. 二群のプロックリスプラウトから抽出した極性代謝産物¹H-NMRスペクトルをもちいた単変量解析。
 (a) volcano plot. 赤で示す領域 ($p < 0.05$ かつ試料間比 > 1.2) は光環境において有意に存在比が大きい代謝産物、一方、青で示す領域 ($p < 0.05$ かつ試料間比 < 0.8) は、暗環境において有意に存在比が大きい代謝産物であることを示す。(b) Box plot. 各試料群における各変量の分布が可視化される。

単変量解析の適用

異なる試料間で得られた変量の間には有意な差があるかを確かめるために、平均値の差の検定が一般的におこなわれます。ここでは、一つの変量の解析に相当するので多変量解析に対して単変量解析と呼びます。ここで、すべての変数に対して単変量解析をおこないました。また、複数の単変量解析の結果を網羅的に可視化するために“volcano plot”を作成しました。

まず、平均値の差の検定をおこなう前に、Shapiro Wilk検定をおこない得られたデータが正規分布に従うとみなせるかを検定しました。正規分布とみなせる変数においては、WelchのT検定を用いて平均値の差の検定をおこないました。一方、正規分布とみなせない変数に関してはMann-WhitneyのU検定を用いて平均値の差の検定をおこないました。複数の変数に対して検定をおこなう場合、変数の数に応じて第一種の過誤が増大します。そこでBenjamini-Hochberg法による補正 (False Discovery Rateによる補正)をおこないました。補正した p 値を用いてvolcanoプロットに用いました。複数のNMRのデータから群間で有意差がある変数を同定する手法の詳細は、参考文献 [7] をご参照ください。

平均値の差の検定からは、それぞれの変数(ここでは代謝産物)において有意な差があるかを確認できますが、すべての変数においてどの変数がどちらの群で有意に多いかを網羅的に把握するのは困難です。Volcano plotを利用することで、どの変数がどちらの群で有意に多いかを網羅的に可視化することができます。Volcano plotは、群間比 (fold change) と p 値 (有意差) の散布図です。散布図上で各プロットの散らばりがよく可視化されるように群間比および p 値を、それぞれ Log_2 および $-\text{Log}_{10}$ で示します。得られる散布図は、視覚的に火山の噴火に似ているため“volcano plot”と呼ばれます。Rの“muma”パッケージでは、各検定、補正した p 値の計算からvolcanoプロットの作成まで自動でおこないます。

Figure 3(a) に、二群のブロッコリスプラウトの極性代謝産物の¹H-NMRスペクトルを用いて作成したvolcano plotを示します。試料間比の閾値には0.8以下 および 1.2以上を、有意差の閾値は $p < 0.05$ を用いました。したがって、 $p < 0.05$ かつ群間比 > 1.2 の領域 (Figure 3(a)の赤ハイライト) にある変数は、光環境において有意に存在比が大きい代謝産物を、他方、 $p < 0.05$ かつ群間比 < 0.8 の領域 (青ハイライト) にある変数は、暗環境において有意に存在比が大きい代謝産物であることを示しています。明環境において有意に多い変数として1.15, 1.19, 4.51 ppmなどの信号 (未同定) が認められ、一方、暗環境において有意に多い変数として7.19, 6.91, 7.95, 7.55, 7.31 ppmなどのHis, Tyr, Phe, Trpなどの芳香族アミノ酸に由来する信号が認められました。

各試料群における各変量の分布は、box plotを作成することで可視化することができます (Figure 3(b))。Box plotでは、各変数における各試料群の最小値、第一四分位点 (25 percentile)、第二四分位点 (50 percentile)、第三四分位点 (75 percentile)、および最大値を示しています。ここで第一四分位点から第三四分位点の間を箱として最小値から第一四分位点および第三四分位点から最大値までをヒゲとして表示しています。例えば Figure 3(b) の左上下段に示す6.91および7.71 ppmの信号 (それぞれTyrおよびTrp由来) のbox plotでは、それぞれの試料群の $n=8$ におけるバケット面積値の分布は、暗環境において高い面積値に分布していることが確認できます。一方、右上下段に示す1.15および2.55 ppmの信号のbox plotでは、明環境において高い面積値に分布していることが確認できます。また、真ん中の上下段に示す3.19および4.27 ppmの信号 (それぞれcholineおよびmalate由来) のvox plotでは、暗環境において若干高い面積値に分布していますが、明環境における分布とおおむね重なっておることが確認できます。

本ノートでは、群情報を目的変数としてデータ解析をおこなう例をご紹介しました。NMR-MPのようにノンターゲットで解析をおこなう場合は、データ取得後の解析が重要であり、群間の比較をしたい場合は、本ノートでお示したように群情報を目的変数にすると効果的です。また、単変量解析も効果的でありvolcano plotなどを使えば変数が多い場合でも、網羅的に可視化できることを示しました。群数が3つ以上になっても同様に解析することが可能です。一方、目的変数が量的変数のときは、回帰分析を用います。

参考文献

- [1] Gentleman; R.C. et al., *Genome Biol.* (2004) **5**, R80. Thévenot; E.A., Roux; A., Xu; Y., Ezan; E., Junot; C., *J. Proteome Res.* (2015) **14**, 3322.
- [2] Gaude; E., et al. *Curr. Metabolomics* (2013) **1**, 180.
- [3] Barker; M. and Rayens; W., *J. Chemometrics* (2003) **17**, 166.
- [4] 弊社アプリケーションノート“仮説なしの多変量解析によるNMRメタボロームデータの要約” (NM170013).
- [5] Eriksson; L. et al., *J. Chemometrics* (2008) **22**, 594.
- [6] Contento; A. L. et al., *Plant Physiol.* (2004) **135**, 2330.
- [7] Goodpaster; A. M. et al., *Anal Biochem.* (2010) **401**, 134.

本誌の記事内容は予告なしに変更することがあります。

Copyright © 2017 JEOL Ltd.

このカタログに掲載した商品は、外国為替及び外国貿易法の安全輸出入管理の規制品に該当する場合がありますので、輸出するとき、または日本国外に持ち出すときは当社までお問い合わせ下さい。



本社・昭島製作所

〒196-8558 東京都昭島市武蔵野3-1-2 TEL: (042) 543-1111(大代表) FAX: (042) 546-3353
www.jeol.co.jp ISO 9001・ISO 14001 認証取得

東京事務所 〒100-0004 東京都千代田区大手町2丁目1番1号 大手町野村ビル13階
営業企画室 TEL: 03-6262-3560 FAX: 03-6262-3577

電子光学機器営業推進室 TEL: 03-6262-3567 分析機器営業推進室 TEL: 03-6262-3568
産業機器営業部 TEL: 03-6262-3570 医用機器ソリューション販売室 TEL: 03-6262-3571

東京支店 〒100-0004 東京都千代田区大手町2丁目1番1号 大手町野村ビル18階 TEL: 03-6262-3580 FAX: 03-6262-3588

電子光学機器営業グループ TEL: 03-6262-3581 分析機器営業グループ TEL: 03-6262-3582

医用機器営業グループ TEL: 03-6262-3583

東京第二事務所 〒190-0012 東京都立川市曙町2丁目8番3号 新鈴巻ビル9階

半導体機器営業室 TEL: 042-528-3491 ソリューションビジネス部 TEL: 042-526-5098

横浜事務所 〒222-0033 神奈川県横浜市港北区新横浜3丁目6番4号 新横浜J&Eビル6階 TEL: 045-474-2181 FAX: 045-474-2180

海外事務所・営業所 Boston, Paris, London, Amsterdam, Stockholm, Sydney, Milan, Singapore, Munich, Beijing, Moscow, Sao Paulo ほか

札幌支店 〒060-0809 北海道札幌市北区北9条西3丁目19番地 ノルテプラザ5階

TEL: 011-726-9680 FAX: 011-717-7305

仙台支店 〒980-0021 宮城県仙台市青葉区中央2丁目2番1号 仙台三菱ビル6階

TEL: 022-222-3324 FAX: 022-265-0202

筑波支店 〒305-0033 茨城県つくば市東新井18番1号

TEL: 029-856-3220 FAX: 029-856-1639

名古屋支店 〒450-0001 愛知県名古屋市中村区那古野1丁目47番1号 名古屋国際センタービル14階

TEL: 052-581-1406 FAX: 052-581-2887

大阪支店 〒532-0011 大阪府大阪市淀川区西中島5丁目14番5号 ニッセイ新大阪南口ビル11階

TEL: 06-6304-3941 FAX: 06-6304-7377

西日本ソリューションセンター

〒532-0011 大阪府大阪市淀川区西中島5丁目14番5号 ニッセイ新大阪南口ビル11階

TEL: 06-6305-0121 FAX: 06-6305-0105

広島支店 〒730-0015 広島県広島市中区橋本町10番6号 広島NSビル5階

TEL: 082-221-2500 FAX: 082-221-3611

高松支店 〒760-0023 香川県高松市海町1-1-12 パシフィックシティ高松5階

TEL: 087-821-0053 FAX: 087-822-0709

福岡支店 〒812-0011 福岡市博多区博多駅前2丁目1番1号 福岡朝日ビル5階

TEL: 092-411-2381 FAX: 092-473-1649