

本研究では、生体試料中の化合物を内因性および外因性に分類し、それらの相互関係を解析するための分析・解析技術の確率を進めた。

分析手法においては、高分解能質量分析計（LC-QToFMS）を用いたノンターゲット分析を基盤とし、SWATH メソッドによる MS/MS スペクトルの取得を行った。分析条件の最適化として、HILIC（親水性相互作用クロマトグラフィー）および C18（逆相）カラムを使用し、ポジティブ/ネガティブイオンモードでの測定を実施することで、極性の異なる幅広い化合物の検出を可能とした。特に、移動相条件の最適化により、内因性および外因性化合物の両方を効率的に分離・検出できる条件を確立した。測定データの QA/QC においては、QC サンプルを用いた分析安定性の継続的モニタリングおよび変動係数 (CV) カットオフ値 (30%) による信頼性の高いピークの選択および、MS/MS までのライブラリとの一致によるアノテーションを実施した。

データ処理とアノテーションに関しては、MS-DIAL ソフトウェアを中心としたデータ処理プラットフォームを構築した。化合物のアノテーションにおいては、MS-DIAL metabolomics MSP、MassBank、HMDB などの複数の公共データベースおよび NIST20 を統合したインハウスライブラリを用いた。また、PubChemLite for Exposomics データベースを用いることで、内因性・外因性化合物の分類手法を確立した。

データ解析手法については、三種類の次元削減手法を比較検討した。具体的には、主成分分析 (PCA)、正則化付き一般化正準相関分析 (rGCCA)、および Uniform Manifold Approximation and Projection (UMAP) を用いた解析を実施した。さらに、OPTICS クラスタリングを用いたサブグループ分析により、化合物プロファイルの違いについて比較した。これらの手法を用いたクラスタリング結果の比較により、PCA と rGCCA が 100%、UMAP は一致率が 92.8%と高い一致率が示され、用

いた手法によらず、試料中に明確な化合物プロファイルの違いが存在することが示唆された (Fig .1)。

クラスタ間で差異が認められた化合物プロファイルについて、外因性化合物と内因性化合物の両面から特徴的な知見が得られた。外因性化合物については、食事由来の化合物として、カフェイン、テオブロミン、パラキサンチンなどのカフェイン関連化合物、環境化学物質としては、フタル酸エステル類、パラベン類などが検出された。内因性化合物の変動については、アミノ酸代謝経路、核酸代謝経路、タンパク質・ミネラル輸送系、エネルギー代謝関連経路、胆汁酸代謝経路など関連する化合物に強度差が認められた。外因性の化合物曝露は生体内の代謝と関連している可能性が示唆されたが、食事由来の化合物と環境化学物質の影響を明確に区別することが困難である点、観察された代謝変動が外因性化合物の直接的な影響であるのか、間接的な影響であるのかの判別が困難である点、さらに一部の化合物については、内因性・外因性の明確な区別が困難である場合がある点などは本研究の限界と言える。

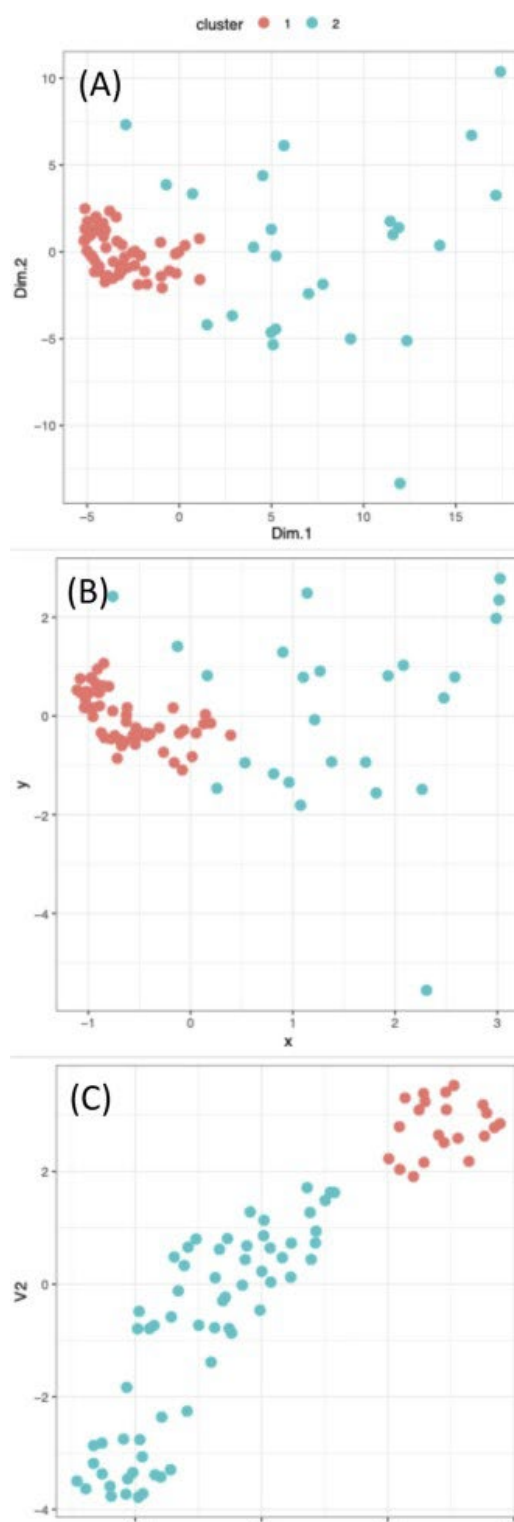


Fig. 1. Score plot for (A) PCA, (B) rGCCA, and (C) UMAP of maternal serum. The green and red points represent the clustering results based on the Optics algorithm.

今後の技術的な展望として、より大規模なコホートデータへの適用、化合物同定アルゴリズムの改良、および生物学的な解釈手法の高度化が必要である。特に、大規模な試料分析を安定して行うための QA/QC は重要であるが、コホート調査に耐えるノンターゲット分析の QA/QC にはバッチ間誤差、日間誤差、装置間誤差など、手法の上で解決すべき課題が多数存在する。

しかし、上記のような問題にも関わらず、本研究で確立された技術基盤は、生体試料のノンターゲット分析における新たな分析・解析手法を提示するものであり、環境化学物質の人体への影響を包括的に評価するための基盤技術となることが期待される。必要がある。

これらに加え、本申請では PBDE 曝露ネコ血清中の lipidome 分析についても実施した。試料 20ul に内部標準及びメタノール/クロロホルム/水を添加し、対象化合物を抽出した。遠心分離の後、上清をバイアルに回収し、LC-QToFMS により分析を行った。これらのデータは現在アノテーションまで完了し、強度の補正について解析を進めている。本試料は複数時点で同一 6 検体から採取された時間を変数にもつ試料であるため、アノテーションおよび強度補正後のデータ解析においても、Multivariate INTEgrative method (MINT) PLS-DA や、timeOMICs などの時系列データについて解析可能なオミクス解析手法を用いる必要がある。また、単純な内部標準による強度補正だけでは QC 試料における繰り返し再現性が十分ではないことから、CV の閾値について検討する必要がある。今回は一旦 CV50 未満の脂質のみをターゲットとし、MINT PLS-DA による解析を行った。解析の結果、大部分の脂質類は曝露群で比較的高値を示し、曝露群、対照群の判別モデルの性能は $AUC = 0.69$ であった (Fig. 2)。本手法においては明確な時系列に伴う成分の変動を検出することはできなかったが、1 日目の測定試料を除いてモデルを検証したところ、 $AUC = 0.75$ となり、曝露から短い時間の試料は判別が難しいことが示唆された。

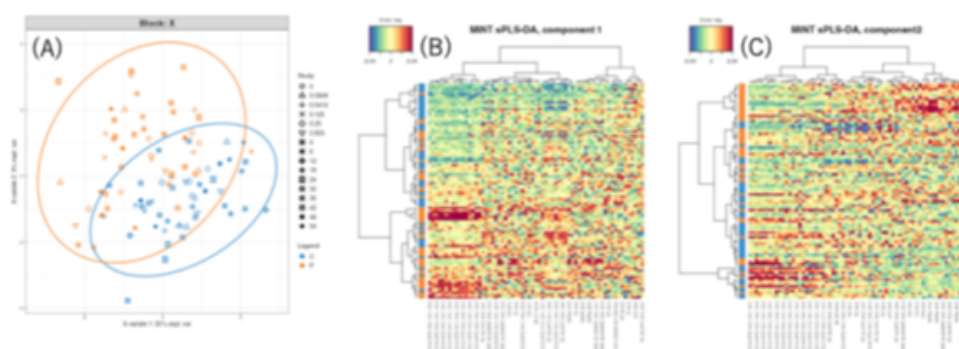


Fig. 2. Score plot for (A) MINT-PLSDA, and heatmap for PC1 (B), and PC2 (C) .

最後にこれら主成分分析を始めとする多変量解析手法および、クラスタ間の比較、データの可視化について、R の解析コードと合わせた共有を行った。主成分分析においては **FactoMineR**, **factoextra** を用いた解析、可視化について共有を行った。特に凡例のグループ分けおよび、解析対象のフィルタリング、測定値の正規化手法に基づく結果の違いに焦点を当て、主成分分析の解釈性向上に関わる内容について主に共有を進めた。

また、多変量のデータにおいて相関関係を可視化する場合に関連し、散布図行列の作成に関する R パッケージの共有を行った。**GGally** パッケージの **ggpairs** 関数を用いることで、10-20 程度の変数について散布図行列を作成する方法について実データを用いて実践した。また、性別などの特徴を用いた層別化を行い、クラスごとの関連の違いに関しても可視化・解析を行った。

最後に濃度依存的のある曝露研究において、単調増加・単調減少などの依存性を解析するための **Jonckheere-Terpstra trend** 検定について共有した。

全体として、古い資料のレシピを使って一定範囲内の解析は可能であることが伺えたものの、実際にコマンドが何をしているのかの理解が十分でないこと、レシピから外れた手法や可視化については実行が難しいことなど、R 導入の敷居の高さが伺えた。書籍などの紹介も行ったが、実際にはある程度の時間を使ってハンズオンでの講習を行うことが望ましい。