**高分辨非靶代谢组学报告**





秉承中科，专“新”致“质”（始于2004年）

**高分辨非靶代谢组学报告**

**Untargeted Metabolomics**

项目名称：

委托单位：

项目编号：

报告时间：

**目录**

**[1.](#_Toc45705867)****[代谢组学概述](#_Toc45705867)** [4](#_Toc45705867)

**[2.](#_Toc45705868)****[项目信息与主要结果概述](#_Toc45705868)** [5](#_Toc45705868)

**[3.](#_Toc45705869)****[项目实验流程](#_Toc45705869)** [6](#_Toc45705869)

**[4.](#_Toc45705870)****[实验仪器和试剂](#_Toc45705870)** [6](#_Toc45705870)

**[5.](#_Toc45705871)****[实验方法](#_Toc45705871)** [7](#_Toc45705871)

**[5.1样品信息](#_Toc45705872)** [7](#_Toc45705872)

**[5.2样本提取方法](#_Toc45705873)** [7](#_Toc45705873)

**[5.3色谱-质谱分析](#_Toc45705874)** [8](#_Toc45705874)

**[5.4 数据分析流程](#_Toc45705875)** [8](#_Toc45705875)

**[5.5 实验数据质量评价](#_Toc45705876)** [9](#_Toc45705876)

**[6.](#_Toc45705877)****[实验结果](#_Toc45705877)** [12](#_Toc45705877)

**[6.1鉴定结果统计与分析](#_Toc45705878)** [12](#_Toc45705878)

**[6.1.1 代谢物鉴定等级](#_Toc45705879)** [12](#_Toc45705879)

**[6.1.2 代谢物鉴定数量统计](#_Toc45705880)** [14](#_Toc45705880)

**[6.1.3 代谢物化学分类归属统计](#_Toc45705881)** [14](#_Toc45705881)

**[6.2 组间差异分析](#_Toc45705882)** [15](#_Toc45705882)

**[6.2.1 单变量统计分析](#_Toc45705883)** [16](#_Toc45705883)

**[6.2.2 多维统计分析](#_Toc45705884)** [18](#_Toc45705884)

**[6.2.3 筛选差异代谢物](#_Toc45705885)** [27](#_Toc45705885)

**[6.3 差异代谢物生物信息学分析](#_Toc45705886)** [29](#_Toc45705886)

**[6.3.1 聚类分析](#_Toc45705887)** [29](#_Toc45705887)

**[6.3.2 相关性分析](#_Toc45705888)** [32](#_Toc45705888)

**[6.3.3 KEGG通路注释与分析](#_Toc45705889)** [37](#_Toc45705889)

**[7.](#_Toc45705890)****[实验结果](#_Toc45705890)** [42](#_Toc45705890)

**[8.](#_Toc45705891)****[附件说明](#_Toc45705891)** [42](#_Toc45705891)

**[9.](#_Toc45705892)****[文献](#_Toc45705892)** [43](#_Toc45705892)

**[10.](#_Toc45705893)****[附录一：实验数据质量评级](#_Toc45705893)** [45](#_Toc45705893)

**[11.](#_Toc45705894)****[附录二：中英文方法（供参考）](#_Toc45705894)** [53](#_Toc45705894)

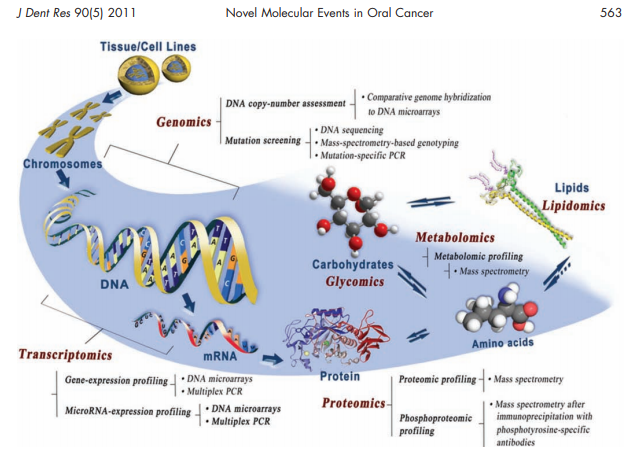
**[12.](#_Toc45705895)****[附录三：后续拓展研究建议（供参考）](#_Toc45705895)** [57](#_Toc45705895)

**[13.](#_Toc45705896)****[附录四：生信数据分析云平台](#_Toc45705896)** [59](#_Toc45705896)

**[14.](#_Toc45705897)****[附录五：组学相关期刊介绍和投稿指南（供参考）](#_Toc45705897)** [60](#_Toc45705897)

1. **代谢组学概述**

代谢组学（metabolomics）诞生于上个世纪末，由英国伦敦[帝国大学](http://zh.wikipedia.org/wiki/%E5%B8%9D%E5%9B%BD%E5%A4%A7%E5%AD%A6" \o "帝国大学)Jeremy Nicholson教授提出(J. et al., 1999)，是研究的是生命体对外界刺激、病理生理变化以及本身[基因突变](http://zh.wikipedia.org/wiki/%E5%9F%BA%E5%9B%A0%E7%AA%81%E5%8F%98" \o "基因突变)而引起的分子质量小于1500Da以内的代谢产物（内源性代谢物）种类、数量及其变化规律的科学。代谢组学是继基因组学、转录组学、蛋白质组学后出现的新兴“组学“，是转录组学和蛋白质组学的延伸，能够更直接、更准确地反映生物体的生理状态，是系统生物学的重要组成部分。系统生物学中，基因组学告诉你可能发生什么，蛋白组学告诉你正在发生什么，而代谢组学是最接近表型的组学，告诉你已经发生了什么。



图片来源于文献：*J Dent Res.*2011 May;90(5):561-72. doi: 10.1177/0022034510383691.

目前代谢组学广泛应用于各研究领域。通过代谢组技术分析实验组与对照组的代谢水平差异，找出差异代谢物，有助于生物标志物的筛选，或研究差异代谢物参与的生物过程（通过代谢通路逆推找出调节酶和基因），揭示其参与的生命活动机制，完成调控通路等方面的研究。因此在疾病诊断、药靶筛选、营养与健康管理、个性化药物治疗、植物生长发育与抗逆等各个研究方向受到越来越多的关注(Nicholson and Lindon, 2008)。

从检测方式上，代谢组学主要分为非靶向分析(untargeted)和靶向分析(targeted)两类。其中，非靶向分析模式通常是建立在高分辨质谱仪上（triple TOF或QE系列等），依赖其强大的高分辨率质量分析器，能够对样本中的各类代谢物进行无偏向、大规模、系统性的检测，提供的是"航拍"的视角，最大程度反映生物体内的代谢水平扰动情况，因此适合项目前期的基础研究；而靶向分析模式通常是建立在单位质量分辨率的质谱仪上（Q-Trap、QQQ等），依赖于三重四极杆串级质谱的高选择能力及高灵敏度，使用针对性开发的样品制备及色谱分离方法，对感兴趣的目标代谢物进行选择性、特异性的检测，适用于特定关注的代谢物检分析及对非靶代谢组筛选到的差异代谢物进行验证。

1. **项目信息与主要结果概述**

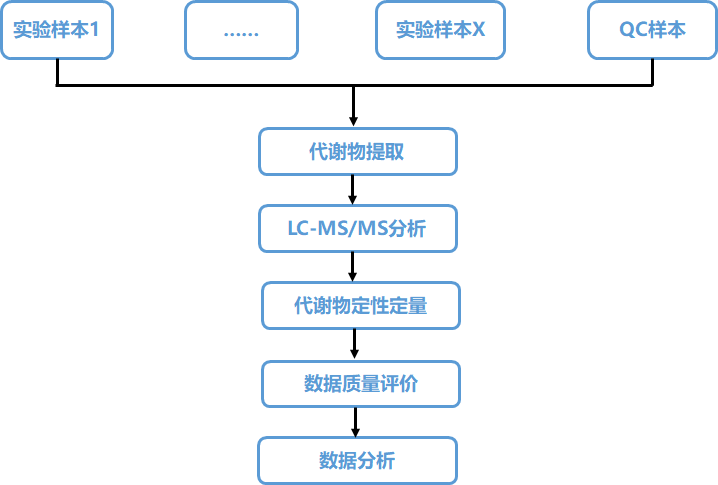
本项目是高分辨非靶向代谢组学分析项目，采用超高效液相色谱-串联飞行时间质谱联用仪（UHPLC-Q-TOF MS）进行检测样本中的代谢物，通过与本地数据库中代谢物的保留时间、分子质量（分子质量误差在<25 ppm内）、二级碎裂谱图、碰撞能等信息进行匹配，对生物样本中的代谢物进行结构鉴定，并对鉴定结果进行严格的人工二次核对、确认。项目信息和主要实验结果总结如表1所示。项目实验流程、实验方法和数据分析结果请查看后续报告内容。

表1 项目信息与主要结果概述

|  |  |  |
| --- | --- | --- |
| **项目信息** | | |
| 样本总数 |  | |
| 总分析对比次数 |  | |
| **主要实验结果** | | |
| 总代谢物鉴定数量 |  | |
| **组间差异对比分析结果** | | |
| 对比组 | 总差异代谢物数量 | 显著性差异代谢通路  （仅展示Top3 KEGG通路富集结果，详见后续实验报告及附件） |
|  |  |  |

1. **项目实验流程**

主要步骤包括：样品制备、QC制备、样品LC-MS/MS质谱分析、数据分析和实验报告等。



代谢组学技术路线

注：QC样本是由待测样本等量混合制成，在待测样本LC-MS/MS进样前、进样中和进样后上机检测。

1. **实验仪器和试剂**

AB Triple TOF 6600质谱仪(AB SCIEX)

Agilent 1290 Infinity LC超高压液相色谱仪(Agilent)

低温高速离心机 (Eppendorf 5430R)

色谱柱: Waters, ACQUITY UPLC BEH Amide 1.7 μm, 2.1 mm× 100 mm column

乙腈(Merck,1499230-935)

乙酸铵(Sigma,70221)

1. **实验方法**

**5.1样品信息**

待测样本信息：样本具体信息见表2。

表2 样品信息

|  |  |  |
| --- | --- | --- |
| **样本分组名称** | **数量** | **样品状态** |
|  |  |  |

**5.2样本提取方法**

样本在4℃环境下缓慢解冻后，取适量样本加入预冷甲醇/乙腈/水溶液（2：2：1，v/v），涡旋混合， 低温超声30min，-20℃静置10 min，14000 g 4℃离心20 min，取上清真空干燥，质谱分析时加入100 μL乙腈水溶液（乙腈：水=1:1，v/v）复溶，涡旋，14000 g 4℃离心15 min，取上清液进样分析。

**5.3色谱-质谱分析**

**（1）色谱条件**

样品采用Agilent 1290 Infinity LC超高效液相色谱系统（UHPLC）HILIC色谱柱进行分离；柱温25℃；流速0.5 mL/min；进样量2 μL；流动相组成A： 水+25 mM乙酸铵+25 mM氨水，B：乙腈；梯度洗脱程序如下：0 --- 0.5 min，95% B ；0.5---7min，B从95%线性变化至65 % ；7---8 min，B从65%线性变化至40%；8--- 9 min，B维持在40% ；9---9.1 min，B从40%线性变化至95%； 9.1--- 12 min，B维持在95%；整个分析过程中样品置于4℃自动进样器中。为避免仪器检测信号波动而造成的影响，采用随机顺序进行样本的连续分析。样本队列中插入QC样品，用于监测和评价系统的稳定性及实验数据的可靠性。

**（2）Q-TOF 质谱条件**

采用AB Triple TOF 6600质谱仪进行样本一级、二级谱图的采集。

HILIC 色谱分离后的ESI源条件如下：Ion Source Gas1（Gas1）：60，Ion Source Gas2（Gas2）：60，Curtain gas （CUR）：30，source temperature：600℃，IonSapary Voltage Floating （ISVF）±5500 V（正负两种模式）；TOF MS scan m/z range：60-1000 Da，product ion scan m/z range：25-1000 Da，TOF MS scan accumulation time 0.20 s/spectra, product ion scan accumulation time 0.05 s/spectra；二级质谱采用information dependent acquisition （IDA）获得，并且采用high sensitivity 模式，Declustering potential（DP）：±60 V（正负两种模式），Collision Energy ：35±15 eV，IDA设置如下 Exclude isotopes within 4 Da，Candidate ions to monitor per cycle：10。

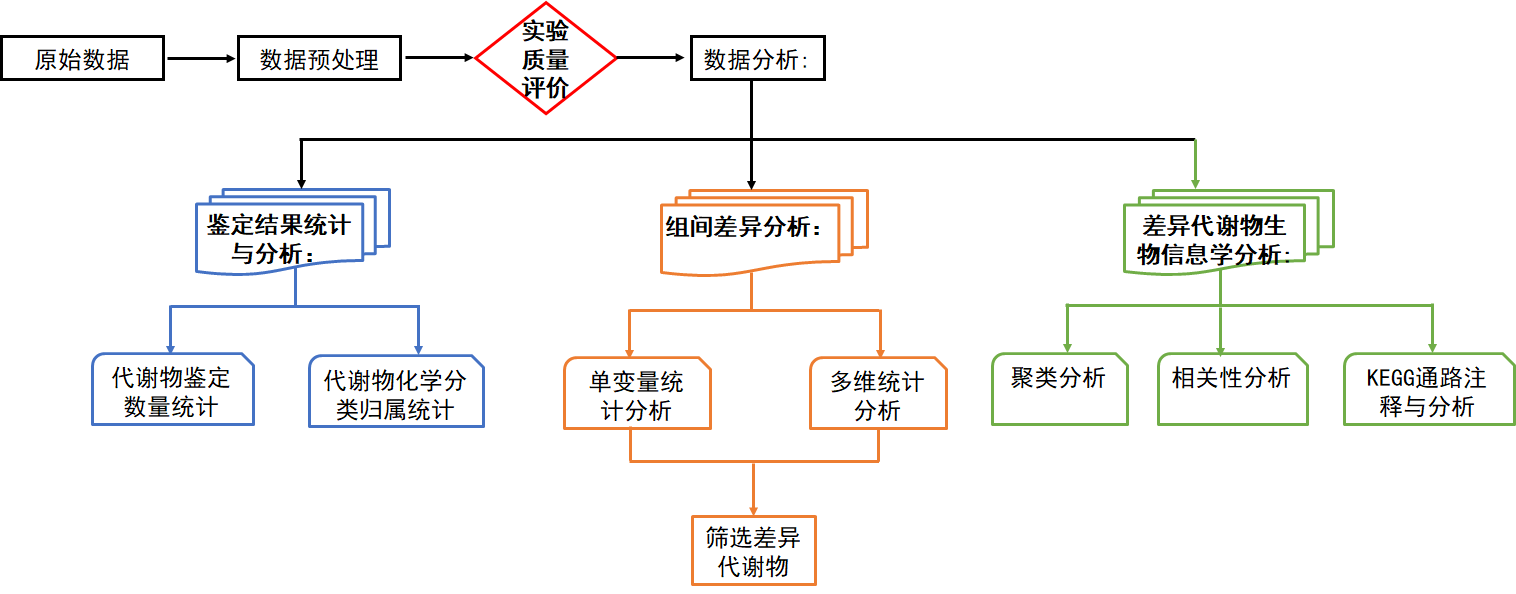
**5.4 数据分析流程**

Wiff格式的原始数据经ProteoWizard转换成.mzXML格式，然后采用XCMS软件进行峰对齐、保留时间校正和提取峰面积。对XCMS提取得到的数据首先进行代谢物结构鉴定、数据预处理，然后进行实验数据质量评价，最后再进行数据分析。



原始数据预处理流程

数据分析内容包括单变量统计分析、多维统计分析、差异代谢物筛选、差异代谢物相关性分析、KEGG通路分析等内容。



数据分析流程

**5.5 实验数据质量评价**

本实验对仪器的稳定性、实验的重复性、数据质量的可靠性进行全面评价。总体来说，本次试验的仪器分析系统稳定性较好，试验数据稳定可靠。在试验中获得的代谢谱差异能反映样本间自身的生物学差异。

通过**六项质控**内容进行评价，部分质控内容：QC样本总离子流图（TIC）的比较和总体样本主成分分析（PCA）评价结果如下所示。

1. **QC样本总离子流图（TIC）的比较**

将QC样本总离子流图（Total ion chromatogram ，TIC）进行谱图重叠比较，如图1a和图1b所示。实验结果表明各色谱峰的响应强度和保留时间基本重叠，说明在整个实验过程中仪器误差引起的变异较小。

**[pos-tic]**

图1a 正离子模式QC样本总离子流图重叠谱图

**[neg-tic]**

图1b 负离子模式QC样本总离子流图重叠谱图

注：图中横坐标表示各色谱峰的保留时间，纵坐标表示峰的强度值。

**2）总体样本主成分分析（PCA）**

将所有实验样本和QC样本提取得到的峰进行PCA分析，见图2a和图2b所示。实验结果表明正、负离子模式下QC样本紧密聚集在一起，说明实验的重复性好。

[qc\_pca\_pos]

图2a 正离子模式总体样本的PCA分析

[qc\_pca\_neg]

图2b 负离子模式总体样本的PCA分析

注：图中t[1]代表主成分1，t[2]代表主成分2，QC样本的聚集程度反映实验的重复性好坏。

其余4项评价内容为：QC样本相关性、总体样本Hotelling‘s T2检验、QC样本的多变量控制图、QC样本的相对标准偏差，评价结果详见**10. 附录一：实验数据质量评价**。

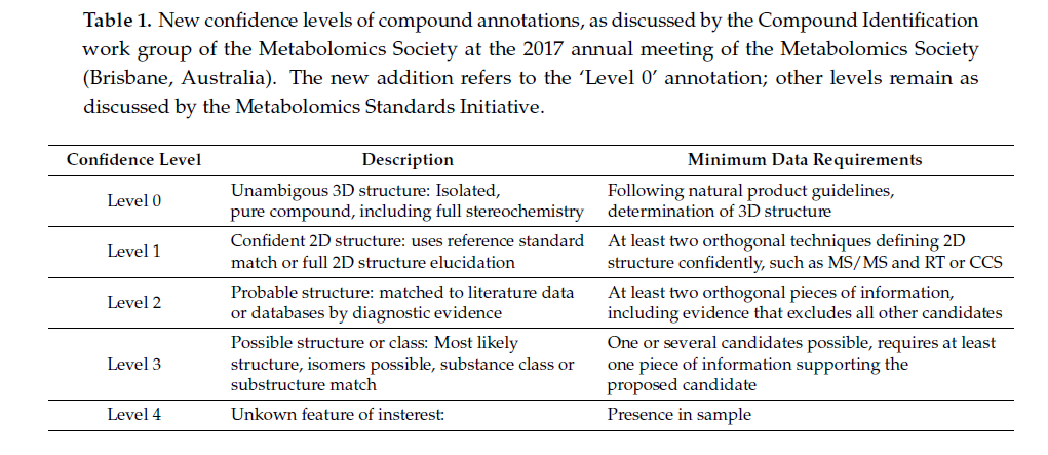
1. **实验结果**

**6.1鉴定结果统计与分析**

**6.1.1 代谢物鉴定等级**

不同于转录组、蛋白组学技术，代谢组学技术发展最晚，研究的代谢物种类繁多，存在大量同分异构体及分子质量接近的代谢物，导致目前代谢物的准确鉴定都是一大难点和挑战。鉴定方式的不同直接影响代谢物鉴定等级的可靠性，进而影响实验结果的可靠性和后续的功能研究，影响重大。国际上代谢组学专家Oliver Fiehn等人建议将代谢物鉴定等级纳入代谢组分析报告中(Blazenovic et al., 2018)。值得注意的是，目前Cell杂志上已经有文章按照Metabolomics Standards Initiative（MSI）的标准提供代谢物鉴定等级了，并且把MSI代谢物等级的定义标准写进了method部分(Contrepois et al., 2020; L et al., 2020)。总的来说，当前审稿人对代谢物鉴定的可靠性要求越来越高，没有经过标准品谱图匹配确认的鉴定结果很容易被拒稿，值得警惕。

一直以来国际上对代谢物的鉴定等级有明确的定义。最早在2007年，MSI（<http://msi-workgroups.sourceforge.net>）的化合物鉴定工作组（Chemical Analysis Working Group）定义了四种代谢物鉴定结果的可靠性等级(Dunn et al., 2012)。后来在2017年澳大利亚布里斯班代谢组学学会年会上，又重新定义了新的代谢物鉴定可靠性等级(Blazenovic et al., 2018)，新增加了”level 0”，从原来的4种，增加到5种，数字越大，可靠性等级越低。如下图所示。



图片来源：*Metabolites.* 2018, *8*(2), 31; <https://doi.org/10.3390/metabo8020031>

**等级0**，具有明确的三维结构和立体化学信息。

**等级1**，可靠的二维结构鉴定，至少需要将真实化学标准品的两个或多个正交性质（如MS/MS谱图、保留时间RT或碰撞截面（CCS）值）与在相同分析条件下分析的感兴趣代谢物的相同性质进行比较。

**等级2级或3级**为假定的注释结果，通常仅基于一个或两个性质，依赖于与不同实验室收集的或用不同分析方法获得的数据进行比较，而不是在相同分析条件下与真实的化学标准进行直接比较。例如，匹配公共数据库, 如HMDB, MoNA, MassBank, METLIN以及NIST等数据库中的MS/MS谱图，代谢物的鉴定等级定义为Level 2。

不能被鉴定为以上等级的为等级4未知物。如Cell上的一篇论文就将MetDNA的鉴定等级定义为Level 4(L et al., 2020)。

本项目采用中科新生命本地自建标准品数据库（in-house database (Shanghai Applied Protein Technology)）搜库(Luo et al., 2017; Zhaobing et al., 2018)。通过与本地数据库中代谢物的保留时间、分子质量（分子质量误差在<25 ppm内）、二级碎裂谱图、碰撞能等信息进行匹配，对生物样本中的代谢物进行结构鉴定，并对鉴定结果进行严格的人工二次核对、确认。**鉴定等级在Level 2以上。**

**6.1.2 代谢物鉴定数量统计**

本项目正负离子模式合并后鉴定total\_meta种代谢物，其中正负离子模式鉴定到的代谢物数量，统计如表3。鉴定结果详见**附件1\_代谢物定性定量结果表.xlsx。**

表3 正负离子模式鉴定到的代谢物数量统计

|  |  |
| --- | --- |
| **检测模式** | **鉴定到的代谢物数量** |
| 正离子模式（Pos） |  |
| 负离子模式（Neg） |  |

**6.1.3 代谢物化学分类归属统计**

本项目鉴定到的所有代谢物（合并正负离子鉴定到的代谢物）根据其化学分类（Chemical Taxonomy）归属信息进行分类统计，各类代谢物数量所占比例如图3所示。具体各代谢物的化学分类归属信息详见**附件1\_代谢物定性定量结果表.xlsx**。

[superclass]

图3 鉴定的代谢物在各化学分类的数量占比

注：图中不同颜色的色块表达不同的化学分类归属条目，百分比代表该化学分类归属条目中的代谢物数量占所有鉴定到的代谢物数量的百分比。没有化学分类归属的代谢物定义为undefined。

输出文件：

Result/ 02. Identified Metabolites\_Stat

**6.2 组间差异分析**

单变量统计分析是从某单一变量水平考察组内变异度和组间差异，而多维统计分析是从总体水平反映组间差异以及反映组内的变异度。代谢组数据具有高维度且变量间高度相关的特点，运用传统的单变量分析无法快速准确地挖掘数据内潜在的信息，所以需要运用多元统计的方法，如PCA、OPLS-DA分析等分析，在最大程度保留原始信息的基础上对采集的多维数据进行降维分析。

同时结合单变量统计分析和多维统计分析结果，是代谢组最常使用的一种组间显著性差异代谢物筛选方法。这些显著性差异代谢物可能是潜在的生物标志物。通过研究差异代谢物参与的代谢通路，逆推找出调节酶和基因，有助于揭示其参与的生命活动机制，完成调控通路等方面的研究。

对正、负离子模式下检测到的两套数据分开分析。以示例对比组为例，对其分析结果在报告中进行展示。其余对比组详见附件。

**6.2.1 单变量统计分析**

单变量统计分析方法是最常用的统计分析方法之一。在进行两组样本间的差异分析时，常用的单变量统计分析方法包括变异倍数分析（Fold Change Analysis，FC Analysis）、T检验/非参检验。

基于单变量分析，对正、负离子模式下检测到的所有代谢物（含未被鉴定的代谢物）进行差异分析。FC > 1.5或FC<0.67，P value < 0.05的差异代谢物，，采用火山图的形式来进行可视化展示，结果如图4所示。

[pos\_volcano]

图4a 正离子模式火山图（颜色与差异代谢物上下调相关）

[neg\_volcano]

图4b 负离子模式火山图（颜色与差异代谢物上下调相关）

注：图中横坐标为差异表达倍数（Fold Change）的log2的对数值，纵坐标为显著性P value的-log10的对数值。显著差异代谢物：满足FC > 1.5，P value < 0.05代谢物用玫红色表示，满足FC<0.67，P value < 0.05的代谢物用蓝色来表示。非显著差异代谢物用黑色来表示。

为了直观便于观察差异代谢物的分类归属，用不同颜色来做区分，结果如图5所示。

[pos\_volcano\_superclass]

图5a 正离子模式火山图（颜色与差异代谢物化学分类相关）

[neg\_volcano\_superclass]

图5b 负离子模式火山图（颜色与差异代谢物化学分类相关）

注：图中横坐标为差异表达倍数（Fold Change）的log2的对数值，纵坐标为显著性P value的-log10的对数值。差异代谢物（FC > 1.5或FC<0.67，P value < 0.05）：按照其化学分类归属信息，用不同的颜色来区分。没有化学分类归属以及未被定性的的差异代谢物定义为undefined。显著差异代谢物用黑色来表示。

输出文件：

Result/ 03. Difference Analysis/Univariate Statistical Analysis【单变量统计分析】

**6.2.2 多维统计分析**

**（1） 主成分分析（PCA）**

主成分分析（Principal Component Analysis, PCA）是一种非监督的数据分析方法，它将原本鉴定到的所有代谢物重新线性组合，形成一组新的综合变量，同时根据所分析的问题从中选取几个综合变量，使它们尽可能多地反映原有变量的信息，从而达到降维的目的。同时，对代谢物进行主成分分析，还能从总体上反映样本组间和组内的变异度。因此在数据分析中，一般先采用PCA方法，观察组间样本的总体分布趋势和组间样本的差异度。

分别对各个比较组做PCA分析，以示例对比组为例，PCA得分图见图6。

[pca\_pos]

图6a 正离子模式PCA得分图

[pca\_neg]

图6b 负离子模式PCA得分图

注：图中t[1]代表主成分1，t[2]代表主成分2，椭圆代表95%置信区间。同一颜色的点表示组内的各个生物学重复，点的分布状态反映出的是组间、组内的差异度。

经7-fold cross-validation（7次循环交互验证）得到的PCA模型参数见表4。R2X越接近1表明模型越稳定可靠。

表4a 正离子模式PCA模型参数

|  |  |  |
| --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** |
|  |  |  |

表4b 负离子模式PCA模型参数

|  |  |  |
| --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** |
|  |  |  |

注：表中A：表示主成分数；R2X：表示模型解释率。

**（2）偏最小二乘判别分析（PLS-DA）**

偏最小二乘判别分析（Partial Least Squares Discrimination Analysis, PLS-DA）是一种有监督的判别分析统计方法。该方法运用偏最小二乘回归建立代谢物表达量与样品类别之间的关系模型，来实现对样品类别的预测。通过建立的判别模型，可以从数据集中筛选出与分组相关的差异脂类物质。

示例对比组的PLS-DA模型得分图见图7。由此可见，PLS-DA模型能区分两组样本。

[plsda\_pos]

图7a 正离子模式PLS-DA得分图

[plsda\_neg]

图7b 负离子模式PLS-DA得分图

注：图中t[1]代表主成分1，t[2]代表主成分2，椭圆代表95%置信区间。同一颜色的点表示组内的各个生物学重复，点的分布状态反映出的是组间、组内的差异度。

经7-fold cross-validation（7次循环交互验证）得到的模型评价参数（R2Y，Q2）列于表5。一般Q2大于0.5，表明模型稳定可靠，0.3<Q2≤0.5，表明模型稳定性较好，Q2<0.3，表明模型可靠性较低。

表5a 正离子模式PLS-DA模型的评价参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** | **R2Y（cum）** | **Q2（cum）** |
|  |  |  |  |  |

表5b 负离子模式PLS-DA模型的评价参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** | **R2Y（cum）** | **Q2（cum）** |
|  |  |  |  |  |

注：表中A：表示主成分数；R2X：表示模型对X变量解释率；R2Y：表示模型对Y变量的解释率；Q2：表示模型预测能力。

为避免有监督模型在建模过程中发生过拟合，采用置换检验（Permutation test）对模型进行检验，以保证模型的有效性。图8显示了示例对比组PLS-DA模型的置换检验图，随着置换保留度逐渐降低，随机模型的R2和Q2均逐渐下降，说明原模型不存在过拟合现象，模型稳健性良好。

[plsda\_permutation\_pos]

图8a 正离子模式PLS-DA置换检验

[plsda\_permutation\_neg]

图8b 负离子模式PLS-DA置换检验

注：图中横坐标表示置换保留度，即与原模型Y变量顺序一致的比例，纵坐标表示R2和Q2的值。绿色的点表示R2，蓝色的点表示Q2，两条虚线分别表示R2和Q2的回归线。右上角的R2和Q2表示置换保留度等于1，即原模型的R2和Q2值。

**（3）正交偏最小二乘判别分析（OPLS-DA）**

正交偏最小二乘判别分析（OPLS-DA）是一种对PLS-DA进行修正的分析方法，可以滤除与分类信息无关的噪音，提高了模型的解析能力和有效性；在OPLS-DA得分图上，有两种主成分，即预测主成分和正交主成分。OPLS-DA将组间差异最大化的反映在t[1]上，所以从t[1]上能直接区分组间变异，而在正交主成分to[1]上则反映了组内的变异。

示例对比组的OPLS-DA模型得分图见图9，可见OPLS-DA模型能区分两组样本。

[oplsda\_pos]

图9a 正离子模式OPLS-DA得分图

[oplsda\_neg]

图9b 负离子模式OPLS-DA得分图

注：图中t[1]代表主成分1，to[1]代表主成分2，椭圆代表95%置信区间。同一颜色的点表示组内的各个生物学重复，点的分布状态反映出的是组间、组内的差异度。

经7-fold cross-validation（7次循环交互验证）得到的模型评价参数（R2Y，Q2）列于表6。一般Q2大于0.5，表明模型稳定可靠，0.3<Q2≤0.5，表明模型稳定性较好，Q2<0.3，表明模型可靠性较低。

表6a 正离子模式OPLS-DA模型的评价参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** | **R2Y（cum）** | **Q2（cum）** |
|  |  |  |  |  |

表6b 负离子模式OPLS-DA模型的评价参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **样品分组** | **A** | **R2X（cum）** | **R2Y（cum）** | **Q2（cum）** |
|  |  |  |  |  |

注：表中A：表示主成分数；R2X：表示模型对X变量解释率；R2Y：表示模型对Y变量的解释率；Q2：表示模型预测能力。

为避免有监督模型在建模过程中发生过拟合，采用置换检验（Permutation test）对模型进行检验，以保证模型的有效性。图10显示了示例对比组OPLS-DA模型的置换检验图，随着置换保留度逐渐降低，随机模型的R2和Q2均逐渐下降，说明原模型不存在过拟合现象，模型稳健性良好。

[oplsda\_permutation\_pos]

图10a 正离子模式OPLS-DA置换检验

[oplsda\_permutation\_neg]

图10b 负离子模式OPLS-DA置换检验

注：图中横坐标表示置换保留度，即与原模型Y变量顺序一致的比例，纵坐标表示R2和Q2的值。绿色的点表示R2，蓝色的点表示Q2，两条虚线分别表示R2和Q2的回归线。右上角的R2和Q2表示置换保留度等于1，即原模型的R2和Q2值。

输出文件：

Result/ 03. Difference Analysis/Multivariate Statistical Analysis【多维统计分析】

**6.2.3 筛选差异代谢物**

OPLS-DA模型得到的变量权重值（Variable Importance for the Projection, VIP）能够用于衡量各代谢物的表达模式对各组样本分类判别的影响强度和解释能力，挖掘具有生物学意义的差异脂质分子。通常VIP>1的代谢物被认为在模型解释中具有显著贡献。

代谢组学通常以严格的OPLS-DA VIP>1和P value < 0.05为显著性差异代谢物筛选标准，本实验以此作为筛选标准。以示例对比组为例，显著性差异代谢物如表7所示。详见**：组间差异代谢物表.xlsx**。

表7a 正离子模式显著性差异代谢物

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **adduct** | **Name** | **VIP** | **FC** | **P-value** | **m/z** | **rt（s）** |
|  |  |  |  |  |  |  |

表7b 负离子模式显著性差异代谢物

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **adduct** | **Name** | **VIP** | **FC** | **P-value** | **m/z** | **rt（s）** |
|  |  |  |  |  |  |  |

注：adduct表示化合物的加合离子信息；Name表示代谢物的名称；VIP表示变量投影重要度，值越大，表示越重要；FC表示差异倍数；P-value表示显著性分析的P值，P值越小，表示差异越显著；m/z表示质荷比；rt（s）表示代谢物在色谱上的保留时间，也即出峰时间，单位为秒。

表中，用黄色背景色表示的是OPLS-DA VIP>1和P value < 0.05的差异代谢物（后续分析仅针对这部分显著性差异代谢物），蓝色背景色表示的是OPLS-DA VIP>1和P value < 0.1的差异代谢物（后续分析不涉及这部分差异代谢物）。若以OPLS-DA VIP>1和P value < 0.05为标准，筛选到的差异代谢物数量较少，可考虑将标准降低，调整为OPLS-DA VIP>1和P value < 0.1。

以柱状图直观展示鉴定到的显著性差异代谢差异倍数变化，如图11所示。

[pos\_fc\_plot]

图11a 正离子模式显著性差异代谢物表达差异倍数分析

[neg\_fc\_plot]

图11b 负离子模式显著性差异代谢物表达差异倍数分析

注：图中横坐标表示差异表达倍数，红色表示差异表达倍数大于1，绿色表示差异表达倍数小于1。纵坐标表示显著性差异代谢物。

输出文件：

Result/ 03. Difference Analysis/ Differential Metabolites【差异代谢物】

**6.3 差异代谢物生物信息学分析**

对筛选到的显著性差异代谢物（**代谢物需同时满足OPLS-DA VIP>1和P value < 0.05，有定性名称**）进行后续的生信分析，包括聚类分析、相关性分析、通路分析等内容。

**6.3.1 聚类分析**

为了更全面直观地显示样本之间的关系以及代谢物在不同样本中的表达模式差异，将所有样本及差异代谢物的表达量进行距离矩阵计算，并用层次聚类(Hierarchical cluster)进行聚类分析。层次聚类首先需要计算样本和样本的距离，或者代谢物与代谢物的距离，每次找到距离最短的两个cluster，然后进行合并成一个大的cluster，直到全部合并为一个cluster。整个过程就是建立一个树结构。

（1）样本层次聚类树

以示例对比组为例，对各组样本进行层次聚类分析，形成表现样本间相似度的聚类树，结果如图12所示。聚在同一簇内的样本，样本之间的相似度更高。

[pos\_hac]

图12a 正离子模式样本层次聚类树

[neg\_hac]

图12b 负离子模式样本层次聚类树

注：图中每列代表一个样品。样本之间相似度相近，聚在同一cluster下。

（2）显著性差异代谢物层次聚类热图

以示例对比组为例，显著性差异代谢物（VIP>1，P value < 0.05）的层次聚类分析结果如图13所示。聚在同一簇内的代谢物具有相似的表达模式，可能具有相似的功能或者共同参与同一代谢过程或者细胞通路。

[pos\_cluster]

图13a 正离子模式显著性差异代谢物层次聚类热图

[neg\_cluster]

图13b 负离子模式显著性差异代谢物层次聚类热图

注：图中每行代表一个差异代谢物（即纵坐标为显著性差异表达的代谢物），每列代表一组样品（即横坐标为样品信息）。红色代表显著性上调，蓝色代表显著性下调，颜色深浅表示上下调的程度，表达模式接近的代谢物聚在左侧同一cluster下。

输出文件：

Result/ 04. Bioinformatics Analysis/Hierarchical Clustering Analysis【聚类分析】

**6.3.2 相关性分析**

相关性分析可以帮助衡量显著性差异代谢物（VIP>1，P value < 0.05）之间的代谢密切程度（metabolic proximities），有利于进一步了解生物状态变化过程中代谢物之间的相互调节关系。具有表达相关性的代谢物，可能共同参与某一生物过程，即功能相关性；此外，不同代谢物之间具有协同或互斥关系，比如某类代谢物变化趋势相同，则为正相关；与某类代谢物变化趋势相反，则为负相关。正相关的代谢物也可能表明其来源于同一合成途径，负相关表明可能被分解用于其他代谢物的合成，即合成转化关系。

基于相关性分析方法，对显著性差异代谢物之间的相关性进行分析，相关性分析结果见**附件Correlation.csv表格**。

以示例对比组为例，相关性分析的结果以相关性热图的形式来进行可视化展示，如图14所示。

[pos\_corr]

图14a 正离子模式相关性热图

[neg\_corr]

图14b 负离子模式相关性热图

注：红色表示正相关，蓝色表示负相关，白色表示非显著性相关。颜色深浅与相关性系数的绝对值大小有关，即正相关或负相关的程度越高，颜色越深。点的大小与相关性的显著性有关，越显著，p值越小，点越大。

为了更直观的揭示各类代谢物之间的共调节关系，相关性矩阵（图14）被转换成和弦图和网络图，如图15和图16所示。和弦图和网络图均展示的是相关性系数|r|>0.8且p<0.05的代谢物分子对(Aviram et al., 2016)，此标准可根据实际的情况调整。和弦图能更好的展示各类代谢物之间的相关性，网络图能更好的展示各种代谢物之间的相关性(Huang et al., 2013)，各有优点。

[pos\_circlize]

图15 a  正离子模式和弦图

[neg\_circlize]

图15b 负离子模式和弦图

注：图中内圈link的起点代表各显著性差异代谢物，外圈上的弧线表示显著性差异代谢物所属分类。彩色线条表示各类代谢物内部的相关性，线条与亚类同色。深灰色线条表示不同类别代谢物之间的相关性。

[pos\_network]

图16a 正离子模式网络图

[neg\_network]

图16b 负离子模式网络图

若需要对网络图进行美化，可参考附件中科新生命《Cytoscape使用指南》进行修改。

注：图中点代表显著性差异代谢物，点的大小与连接度degree相关，degree越大，点越大。线条的颜色代表相关性，红色表示正相关，蓝色表示负相关。线条的粗细代表相关性系数绝对值的大小，线条越粗，相关性越大。

输出文件：

Result/ 04. Bioinformatics Analysis/ Correlation Analysis【聚类分析】

**6.3.3** **KEGG通路注释与分析**

KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.kegg.jp/)是常用于通路研究的数据库之一，它是由研究人员阅读海量文献以特定的图形语言描述众多的代谢途径以及各途径之间的相互关系。KEGG数据库收录了新陈代谢，遗传信息加工，环境信息加工，细胞过程，生物体系统，人类疾病以及药物开发等多个方面的通路信息，KEGG数据库相关资料参见: <http://www.genome.jp/kegg/pathway.html>。在生物体内，不同代谢物相互协调行使其生物学功能，基于KEGG Pathway的分析有助于更进一步了解其生物学功能。

**KEGG通路注释与分析前，先对正负离子模式筛选到的差异代谢物合并。**以示例对比组为例，分析结果如下所示：

**6.3.3.1 KEGG通路注释结果**

显著性差异代谢物（VIP>1，P value < 0.05）的KEGG通路注释结果如表8和表9所示，详见**附件: Kegg.xlsx**。

表8 差异物的KEGG通路注释结果（仅展示表头）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metabolite** | **KEGG.ID** | **cpdName** | **Map\_ID** | **Map\_Name** | **URL** |

注：Metabolite表示显著性差异代谢物名称；KEGG.ID表示代谢物的KEGG ID号；cpdName表示该代谢物在KEGG数据库中的名称；Map\_ID表示代谢通路的ID号（（备注：同一代谢物可能会参与多条代谢途径，被注释到多条代谢通路）；Map\_Name表示代谢通路的名称；URL表示该条代谢通路的网页版查看链接。

表9 KEGG代谢通路中的差异代谢物（仅展示表头）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Pathway\_Hierarchy1** | **Pathway\_Hierarchy2** | **Map\_ID** | **Map\_Name** | **cpdName** | **KEGG.ID** | **Num\_cpd** | **URL** |

注：Pathway\_Hierarchy1和Pathway\_Hierarchy2分别表示代谢通路的第一层和第二层归属；Map\_ID、Map\_Name、cpdName、KEGG.ID和URL的解释同上（表7）；Num\_cpd表示注释到该条通路的差异代谢物数量。

输出文件：

Result/ 04. Bioinformatics Analysis/ KEGG Analysis/Kegg.xlsx

**6.3.3.2 KEGG通路图**

进一步通过KEGG pathway mapper功能对差异代谢通路进行展示，并根据上下调信息对差异代谢物着色显示。本项目选择其中一条通路图进行展示，如图16所示。其余详见**附件KEGG map文件夹**。

[pathway]

图17 KEGG通路图

注：代谢通路图中的小圆圈代表代谢物。通路图中红色标识的代谢物为实验检测到的差异代谢物。

为了便于观察KEGG代谢通路中注释到的各差异代谢物的表达情况，本项目选择差异代谢物数量**大于5**的KEGG代谢通路，以热图的形式对KEGG代谢通路中的差异代谢物进行展示，如图17所示。每条差异代谢通路对应一张热图，详见**附件KEGG Pathway\_Metabolites Heatmap文件夹**。

[kegg\_heatmap]

图18 KEGG通路的差异代谢物聚类热图

注：图中每行代表一个差异代谢物（即纵坐标为显著性差异表达的代谢物），每列代表一组样品（即横坐标为样品信息）。红色代表显著性上调，蓝色代表显著性下调，颜色深浅表示上下调的程度，表达模式接近的代谢物聚在左侧同一cluster下。

输出文件：

Result/ 04. Bioinformatics Analysis/ KEGG Analysis/ KEGG map【KEGG通路图】

Result/ 04. Bioinformatics Analysis/ KEGG Analysis/ KEGG Pathway\_Metabolites Heatmap【KEGG通路代谢物聚类热图】

**6.3.3.3 KEGG通路富集图**

KEGG通路富集分析是以KEGG通路为单位，以该物种或亲缘关系较接近的物种所参与的代谢通路为背景，通过Fisher精确检验（Fisher’s Exact Test），来分析计算各个通路代谢物富集度的显著性水平，从而确定受到显著影响的代谢和信号转导途径。P值越小，则该代谢通路的差异性越显著。

代谢通路富集分析的结果以气泡图和柱状图两种形式进行展示，结果如图19a和图19b所示。

[kegg\_enrich1]

图19a KEGG富集通路图（气泡图）

注：气泡图中每一个气泡代表一个代谢通路（根据P value选择显著性最高的前20条），气泡所在横坐标和气泡大小表示该通路在拓扑分析中的影响因子大小，大小越大影响因子越大；气泡所在纵坐标和气泡颜色表示富集分析的P 值（取负常用对数，即-log10 P-value），颜色越深P 值越小，富集程度越显著。

[kegg\_enrich2]

图19a KEGG富集通路图（柱状图）

注：柱状图中纵轴代表各KEGG代谢通路，横轴表示各KEGG代谢通路中包含的差异表达代谢物数目。颜色表示富集分析的P 值，颜色越深P 值越小，富集程度越显著。

输出文件：

Result/ 04. Bioinformatics Analysis/ KEGG Analysis/ KEGG Enrichment Analysis【KEGG通路富集分析】

**6.3.3.4 KEGG代谢通路整体变化分析**

Differential Abundance Score(差异丰度得分)是一种基于通路的代谢变化分析方法，差异丰度得分可以捕捉到某一途径中所有代谢物的平均、总体变化。

以示例对比组为例，所有差异代谢通路的差异丰度得分如图20a和图20b所示。对于该分析方法的介绍，详见参考文献(Hakimi et al., 2016)。

[kegg\_dascore1]

图20a 所有差异代谢通路的差异丰度得分图

所有的差异代谢通路按其上一层级Pathway\_Hierarchy进行分类归属，然后重新作图展示，结果如图20b所示。

[kegg\_dascore2]

图20b 所有差异代谢通路的差异丰度得分图（按照Pathway\_Hierarchy进行分类归属）

注：图中Y轴表示差异通路的名称，X轴坐标表示差异丰度得分 (DA score)。DA score为代谢途径中所有代谢物的整体总变化。得分1表示该通路中所有鉴定到的代谢物表达趋势上调,-1该通路中所有鉴定到的代谢物表达趋势下调。线段的长度表示DA score的绝对值，线段端点的圆点大小表示该通路中代谢物的数目多少，点越大表示代谢物数目越多。线段和圆点颜色的深浅和DA score值成比例，红色越深，表示该通路整体表达情况越倾向于上调，蓝色越深，表示该通路整体表达情况越倾向于下调。

输出文件：

Result/ 04. Bioinformatics Analysis/ KEGG Analysis/ Differential Abundance Score【KEGG代谢通路整体变化分析】

1. **实验结果**

本实验采用基于UHPLC-Q-TOF MS技术的代谢组学方法分别对样本进行了代谢轮廓变化分析。质量控制实验表明，本次实验的仪器分析系统稳定性较好，试验数据稳定可靠。在实验中获得的代谢谱差异能反映样本间自身的生物学差异。

1. **附件说明**
2. **附件1 \_代谢物定性定量结果表.xlsx**
3. **附件2 数据结果相关文件：**

**Result/ 01. QC【实验质量控制】**

**Result/ 02. Identified Metabolites\_Stat【鉴定结果统计与分析】**

**Result/ 03. Difference Analysis【组间差异分析】**

----Univariate Statistical Analysis【单变量统计分析】

----Multivariate Statistical Analysis【多维统计分析】

----Differential Metabolites【差异代谢物】

**Result/ 04. Bioinformatics Analysis【差异代谢物生信分析】**

----Hierarchical Clustering Analysis【聚类分析】

----Correlation Analysis【相关性分析】

----KEGG Analysis【KEGG分析】

--- Kegg.xlsx【KEGG注释结果】

----KEGG Map【KEGG通路图】

----KEGG Pathway\_Metabolites Heatmap【KEGG通路代谢物聚类热图】

----KEGG Enrichment Analysis【KEGG通路富集分析】

----Differential Abundance Score【KEGG代谢通路整体变化分析】

1. **文献**

Aviram, R., Manella, G., Kopelman, N.M., Neufeldcohen, A., Zwighaft, Z., Elimelech, M., Adamovich, Y., Golik, M., Wang, C., and Han, X. (2016). Lipidomics Analyses Reveal Temporal and Spatial Lipid Organization and Uncover Daily Oscillations in Intracellular Organelles. Molecular Cell *62*, 636-648.

Blazenovic, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. Metabolites *8*.

Contrepois, K., Wu, S., Moneghetti, K.J., Hornburg, D., Ahadi, S., Tsai, M.S., Metwally, A.A., Wei, E., Lee-McMullen, B., Quijada, J.V.*, et al.* (2020). Molecular Choreography of Acute Exercise. Cell *181*, 1112-1130 e1116.

Dunn, W.B., Erban, A., Weber, R.J.M., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J.*, et al.* (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics *9*, 44-66.

Hakimi, A.A., Reznik, E., Lee, C.-H., Creighton, C., and Hsieh, J. (2016). An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. Cancer Cell *29*, 104-116.

Huang, Q., Tan, Y., Yin, P., Ye, G., Gao, P., Lu, X., Wang, H., and Xu, G. (2013). Metabolic Characterization of Hepatocellular Carcinoma Using Nontargeted Tissue Metabolomics. Cancer Research *73*, 4992-5002.

J., K., Nicholson, J., C., Lindon, E., and Holmes (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica.

L, L., MH, R., B, P., X, S., S, C., H, R., JK, S., R, T., L, S., NC, L.*, et al.* (2020). Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. Cell *181*, 1680-1692.e1615.

Luo, D., Deng, T., Yuan, W., Deng, H., and Jin, M. (2017). Plasma metabolomic study in Chinese patients with wet age-related macular degeneration. Bmc Ophthalmology *17*, 165.

Nicholson, J.K., and Lindon, J.C. (2008). Systems biology: Metabonomics. Nature *455*, 1054-1056.

Shen, X., Gong, X., Cai, Y., Guo, Y., Tu, J., Li, H., Zhang, T., Wang, J., Xue, F., and Zhu, Z. (2016). Normalization and integration of large-scale metabolomics data using support vector regression. Metabolomics *12*, 89.

Siskos, A.P., Jain, P., Romischmargl, W., Bennett, M.H., Achaintre, D., Asad, Y., Marney, L.C., Richardson, L., Koulman, A., and Griffin, J.L. (2017). Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. Analytical Chemistry *89*, 656-665.

Zhaobing, Gu, Lin, Li, Shoukun, Tang, Chuanbin, Liu, Xianhai, and Fu (2018). Metabolomics Reveals that Crossbred Dairy Buffaloes are more Thermotolerant than Holstein cows under Chronic Heat Stress. Journal of agricultural and food chemistry.

1. **附录一：实验数据质量评级**

**1）QC样本总离子流图（TIC）的比较**

将QC样本总离子流图（Total ion chromatogram ，TIC）进行谱图重叠比较，如图1a和图1b所示。实验结果表明各色谱峰的响应强度和保留时间基本重叠，说明在整个实验过程中仪器误差引起的变异较小。

**[pos-tic]**

图1a 正离子模式QC样本总离子流图重叠谱图

**[neg-tic]**

图1b 负离子模式QC样本总离子流图重叠谱图

注：图中横坐标表示各色谱峰的保留时间，纵坐标表示峰的强度值。

**2）总体样本主成分分析（PCA）**

将所有实验样本和QC样本提取得到的峰进行PCA分析，见图2a和图2b所示。实验结果表明正、负离子模式下QC样本紧密聚集在一起，说明实验的重复性好。

[qc\_pca\_pos]

图2a 正离子模式总体样本的PCA分析

[qc\_pca\_neg]

图2b 负离子模式总体样本的PCA分析

注：图中t[1]代表主成分1，t[2]代表主成分2，QC样本的聚集程度反映实验的重复性好坏。

**3）QC样本相关性**

对QC样本进行Pearson相关性分析，如图3a和图3b所示。一般相关性系数大于0.9表明相关性较好。实验结果表明QC样本间的相关性系数都在0.9以上，说明实验重复性较好。

[pos\_scatter]

图3a 正离子模式QC样本相关性图谱

[neg\_scatter]

图3b 负离子模式QC样本相关性图谱

注：图中横、纵坐标代表各QC样本。每个小格中的点代表QC样本提取到的离子峰（代谢物），横坐标和纵坐标代表的是离子峰信号强度值的对数值。

**4）总体样本Hotelling‘s T2检验**

Hotelling’s T2检验通过多元变量建模对样本进行检验，定义了95%或99%置信区间，可用于离群样本的诊断。Hotelling’s T2检验结果见图4。实验结果表明QC样本均在99%置信区间内(Siskos et al., 2017)，说明实验的重复性好。

[pos\_ht2]

图4a 正离子模式总体样本的Hotellings T2 图

[neg\_ht2]

图4b 负离子模式总体样本的Hotellings T2 图

注：图中横坐标代表所有实验样本和QC样本，纵坐标反映置信区间，红色的线定义了99%的置信区间范围。

**5）QC样本的多变量控制图**

多变量控制图（Multivariate Control Chart，MCC）是基于QC样本检测到的离子峰建立的多元变量统计学模型，是用于监控和判断仪器状态是否稳定的一种质量管理工具。

多变量控制图中的每个点代表一个QC样本，X轴是所有QC样本的上机顺序。 由于仪器状态的波动，图中的点呈现上下波动的情况。一般正负3个标准差范围内为正常范围。本项目QC样本的多变量控制图见图5。实验结果表明QC样本的波动都在正负3个标准差范围内，反映仪器的波动在正常范围内，数据可用于后续分析。

**[pos\_mcc]**

图5a 正离子模式QC样本MCC图

[neg\_mcc]

图5b 正离子模式QC样本MCC图

注：图中横坐标代表各QC样本，纵坐标反映标准差，黄色和红色的线分别定义了正负2个、3个标准差范围。

**6） QC样本的相对标准偏差（RSD）**

QC样本离子峰丰度的相对标准偏差（RSD）越小，表明仪器的稳定性越好，是反映数据质量好坏的一个重要指标。本实验QC样本中RSD≤30%的Peak数目占QC样本总Peak数目的比例在80%以上，见图6，表明仪器分析系统稳定性较好，数据可以用于后续分析(Shen et al., 2016)。

[pos\_rsd]

图6a 正离子模式QC样本的相对标准偏差

[neg\_rsd]

图6b 正离子模式QC样本的相对标准偏差

输出文件：

Result/01. QC

1. **附录二：中英文方法（供参考）**

* **Chemicals**

Ammonium acetate (NH4AC), ammonium hydroxide (NH4OH), ammonium fluoride (NH4F), and formic acid (FA) were purchased from Sigma Aldrich. Acetonitrile was purchased from Merck.

* **Sample Collection and Preparation**

**Biofluid Samples (Plasma , Serum , and Urine)**:

Fasting blood samples were collected in 5 mL Vacutainer tubes containing the chelating agent ethylene diamine tetraacetic acid (EDTA), then the samples were centrifuged for 15 min (1500g, 4 °C). Each aliquot (150 μL) of the plasma sample was stored at –80 °C until UPLC-Q-TOF/MS analysis. The plasma samples were thawed at 4 °C and 100 μL aliquots were mixed with 400 μL of cold methanol/acetonitrile (1:1, v/v) to remove the protein. The mixture was centrifuged for 15 min (14000g, 4 °C). The supernatant was dried in a vacuum centrifuge. For LC-MS analysis, the samples were re-dissolved in 100 μL acetonitrile/water (1:1, v/v) solvent.

**Animal Tissues:**

The animal tissues (e.g., mice liver) were quickly frozen in liquid nitrogen immediately after dissection. Then the tissue were cut on dry ice (~10 mg) into an Eppendorf tube (2 mL). The tissue samples with 200 μL of H2O and five ceramic beads were homogenized using the homogenizer. 800 μL methanol/acetonitrile (1:1, v/v) were added to homogenized solution for metabolite extraction. The mixture was centrifuged for 15 min (14000g, 4 °C). The supernatant was dried in a vacuum centrifuge. For LC-MS analysis, the samples were re-dissolved in 100 μL acetonitrile/water (1:1, v/v) solvent.

**Plant Tissues:**

The plant tissues（80mg leaves, or flower tissue） were quickly frozen in liquid nitrogen immediately and grounded into fine powder with a mortar and pestle.1000 μL methanol/acetonitrile /H2O (2:2:1, v/v/v) were added to homogenized solution for metabolite extraction. The mixture was centrifuged for 15 min (14000g, 4 °C). The supernatant was dried in a vacuum centrifuge. For LC-MS analysis, the samples were re-dissolved in 100 μL acetonitrile/water (1:1, v/v) solvent.

**Cultured Cells (Human cells, Hela cells):**

The culture medium from the cultured Hela cells (~2 X 106 cells per sample) was removed using pipette. Then the cells were washed with PBS under 37°C for three times and the PBS was removed. The sample was proceeded with trypsinization for 2 min, then serum medium was added to stop the reaction. The cells were collected into a new centrifuge tube, and centrifuged at 14000g for 5 min to remove supernatant. The cell pellets were washed with PBS under 4°C for three times, and centrifuged at 14000g for 5 min to remove the PBS. Then the cell pellets were quickly frozen in liquid nitrogen and immediately proceeded with extraction or store the material at -80 °C. 1000 μL methanol/acetonitrile /H2O (2:2:1, v/v/v) were added to homogenized solution for metabolite extraction. The mixture was centrifuged for 15 min (14000g, 4 °C). The supernatant was dried in a vacuum centrifuge. For LC-MS analysis, the samples were re-dissolved in 100 μL acetonitrile/water (1:1, v/v) solvent.

To monitor the stability and repeatability of instrument analysis, quality control (QC) samples were prepared by pooling 10 μL of each sample and analyzed together with the other samples. The QC samples were inserted regularly and analyzed in every 5 samples.

* **LC-MS/MS Analysis**

Analyses were performed using an UHPLC (1290 Infinity LC, Agilent Technologies) coupled to a quadrupole time-of-flight (AB Sciex TripleTOF 6600) in Shanghai Applied Protein Technology Co., Ltd.

For HILIC separation, samples were analyzed using a 2.1 mm × 100 mm ACQUIY UPLC BEH 1.7 µm column (waters, Ireland). In both ESI positive and negative modes, the mobile phase contained A=25 mM ammonium acetate and 25 mM ammonium hydroxide in water and B= acetonitrile. The gradient was 85% B for 1 min and was linearly reduced to 65% in 11 min, and then was reduced to 40% in 0.1 min and kept for 4 min, and then increased to 85% in 0.1 min, with a 5 min re-equilibration period employed.

For RPLC separation, a 2.1 mm × 100 mm ACQUIY UPLC HSS T3 1.8 µm column (waters, Ireland) was used. In ESI positive mode, the mobile phase contained A= water with 0.1% formic acid and B= acetonitrile with 0.1% formic acid; and in ESI negative mode, the mobile phase contained A=0.5 mM ammonium fluoride in water and B= acetonitrile. The gradient was 1%B for 1.5 min and was linearly increased to 99% in 11.5 min and kept for 3.5 min. Then it was reduced to 1% in 0.1 min and a 3.4 min of re-equilibration period was employed. The gradients were at a flow rate of 0.3 mL/min, and the column temperatures were kept constant at 25℃. A 2 µL aliquot of each sample was injected.

The ESI source conditions were set as follows: Ion Source Gas1 (Gas1) as 60, Ion Source Gas2 (Gas2) as 60, curtain gas (CUR) as 30, source temperature: 600℃, IonSpray Voltage Floating (ISVF) ± 5500 V. In MS only acquisition, the instrument was set to acquire over the m/z range 60-1000 Da, and the accumulation time for TOF MS scan was set at 0.20 s/spectra. In auto MS/MS acquisition, the instrument was set to acquire over the m/z range 25-1000 Da, and the accumulation time for product ion scan was set at 0.05 s/spectra. The product ion scan is acquired using information dependent acquisition (IDA) with high sensitivity mode selected. The parameters were set as follows: the collision energy (CE) was fixed at 35 V with ± 15 eV; declustering potential (DP), 60 V (+) and −60 V (−); exclude isotopes within 4 Da, candidate ions to monitor per cycle: 10.

* **Data processing**

The raw MS data (wiff.scan files) were converted to MzXML files using ProteoWizard MSConvert before importing into freely available XCMS software. For peak picking, the following parameters were used: centWave m/z = 25 ppm, peakwidth = c (10, 60), prefilter = c (10, 100). For peak grouping, bw = 5, mzwid = 0.025, minfrac = 0.5 were used. CAMERA (Collection of Algorithms of MEtabolite pRofile Annotation) was sued for annotation of isotopes and adducts. In the extracted ion features, only the variables having more than 50% of the nonzero measurement values in at least one group were kept. Compound identification of metabolites was performed by comparing of accuracy m/z value (<25 ppm), and MS/MS spectra with an in-house database established with available authentic standards.

* **Statistical analysis**

After normalized to total peak intensity, the processed data were analyzed by R package (ropls), where it was subjected to multivariate data analysis, including Pareto-scaled principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA). The 7-fold cross-validation and response permutation testing were used to evaluate the robustness of the model. The variable importance in the projection (VIP) value of each variable in the OPLS-DA model was calculated to indicate its contribution to the classification. Metabolites with the VIP value >1 was further applied to Student’s t-test at univariate level to measure the significance of each metabolite, the p values less than 0.05 were considered as statistically significant.

1. **附录三：后续拓展研究建议（供参考）**

基于目前的方案设计，本研究还可考虑的扩展分析内容与方向，提供建议如下，供您参考：

* 非靶代谢组学（偏极性代谢物）+脂质组学，描绘代谢变化全景

生物脂质，包括甘油酯、甘油磷脂、鞘脂、糖脂等，既是能量物质，也是生物膜(亚细胞)的结构与功能基础,以及重要的信号小分子和能量物质。因此，脂质不仅参与生长发育、神经信号转导、光合作用等多种生理过程，而且脂质代谢紊乱还与各种病理过程以及植物物胁迫等密切相关。

代谢物极性跨度大，很难通过一种体系覆盖所有类型的代谢物。非靶向代谢组学偏向于极性代谢物的检测。脂质组学偏向于非极性的脂类检测。因此非靶向代谢组学（偏极性代谢物）+脂质组学，可以对生物样本内的代谢物进行更全的检测，全面描绘生物体的代谢变化全景图。

* 打通表型与机制研究，探索代谢调控模型

代谢组学反映的是代谢层面的变化。引起代谢变化的调控机制比较复杂，涉及转录、蛋白表达及蛋白修饰，因此多组学之间的联合可以（1）系统描绘从基因/蛋白变化到代谢改变的整个代谢调控过程；（2）基于已知代谢通路，确认基因/蛋白对特定代谢过程的调控作用，组学结果相互验证；（3）依据表达量相关性，挖掘基因/蛋白与代谢物的新的调控关系。

|  |  |
| --- | --- |
| 转录组+代谢组 | 描绘mRNA-基础代谢的调控过程与关系 |
| 蛋白组+代谢组 | 描绘蛋白质-基础代谢的调控过程与关系 |
| 修饰组+代谢组 | 描绘蛋白修饰-基础代谢的调控过程与关系 |

* 是否还关注特定种类的代谢物？

非靶向代谢组学是一种无目标性的大规模检测手段，一次可检测到上千种的代谢物离子峰，缺点是不能保证目标物能够被检测到。如果还特别关注某些特定种类的代谢物，建议非靶代谢组学与靶向代谢组学联合使用。

|  |  |
| --- | --- |
| 靶向目标代谢物分析 | ■ P200 ■ 能量代谢类  ■ 短链脂肪酸类 ■ 中长链脂肪酸类  ■ 胆汁酸类 ■ 氨基酸类  ■ 神经递质类 ■ 花生四烯酸代谢类  ■ 水溶性维生素类 ■ 脂溶性维生素类  ■ 植物激素类，等 |
| 个性化建方法 | 选择性的对某些代谢物进行靶向验证 |

1. **附录四：****生信数据分析云平台**

* APT-BioCloud

对于以上报告的分析内容，若您仍需对分析图片的字体、样式等做调整（或数据修改后需重新分析），您也可以登录中科新生命APT-BioCloud分析云平台，进行一站式自助数据分析。将数据导入至云平台后，您可自行进行统计学、火山图、聚类分析、富集分析等多种生物信息学分析，快速便捷输出个性化的生信分析图。注册与登录地址如下，如有疑问可联系中科新生命当地销售。

注册网址：[http//cloud.aptbiotech.com/#/register](file:///C:\\Users\\wjxu\\Documents\\WeChat%20Files\\wenjiaxu1018\\FileStorage\\File\\2020-05\\http\\cloud.aptbiotech.com\\" \l "/register) 登录网址：[http//cloud.aptbiotech.com/#/login](file:///C:\\Users\\wjxu\\Documents\\WeChat%20Files\\wenjiaxu1018\\FileStorage\\File\\2020-05\\http\\cloud.aptbiotech.com\\" \l "/login)



1. **附录五：组学相关期刊介绍和投稿指南（供参考）**

* **第一档期刊（影响因子9 分以上）**

1. 组学期刊：Nature Genetics，Genome Research， Microbiome，Molecular Systems Biology等
2. 综合性期刊： Cell，Nature，Science，Nature Communications，PNAS，Cell Metabolism，Molecular Cell，Cell Research等
3. 医学专业期刊：Nature Medicine，Cancer Cell，Cancer Research，European Heart Journal，Circulation，Journal of the American College of Cardiology，Gut，Gastroenterology，Hepatology，Diabetes Care，Molecular Psychiatry，Immunity 等
4. 农林专业期刊：Nature Plants，Molecular Plant，Plant Cell等

**组学思路**：在这类高水平期刊中，组学驱动的研究主要包括：1.大队列研究，筛选临床生物标志物以及进行疾病分子分型等。2. 组学+后续功能机制研究。3. 结合时下的热点，如肠道微生物进行相关研究，阐明肠道微生物与疾病的相关性及相互作用机制。

* **第二档期刊（影响因子4~9分）**

1. 组学期刊：Molecular & Cellular Proteomics，Metabolism，Metallomics，Gigascience，Proteomics & Bioinformatics等
2. 综合期刊： Scientific Data， Scientific Reports，Cell Reports，Progress in Lipid Research，Frontiers in Microbiology，mSystems等
3. 医学专业期刊：Advances in Cancer Research，International Journal of Cancer，Diabetologia，Thyroid，Diabetes，Oncogene，mBio等
4. 农林领域专业期刊：New Phytologist，Plant Biotechnology Journal，Plant Physiology，Plant Cell and Environment，Food Chemistry，Journal of Experimental Botany等

**组学思路**：发表在该分数段期刊的组学文章，1. 对于蛋白组学，修饰组学而言，要求组学+功能验证与机制研究（如表达量验证，基因敲除，点突变，位点特异性抗体验证，蛋白活性验证，后续功能实验，动物模型等）。2. 对于脂质组学相对例外，研究基础浅，新颖性高，纯脂质组学也能发表文章，对后续验证要求相对较低。

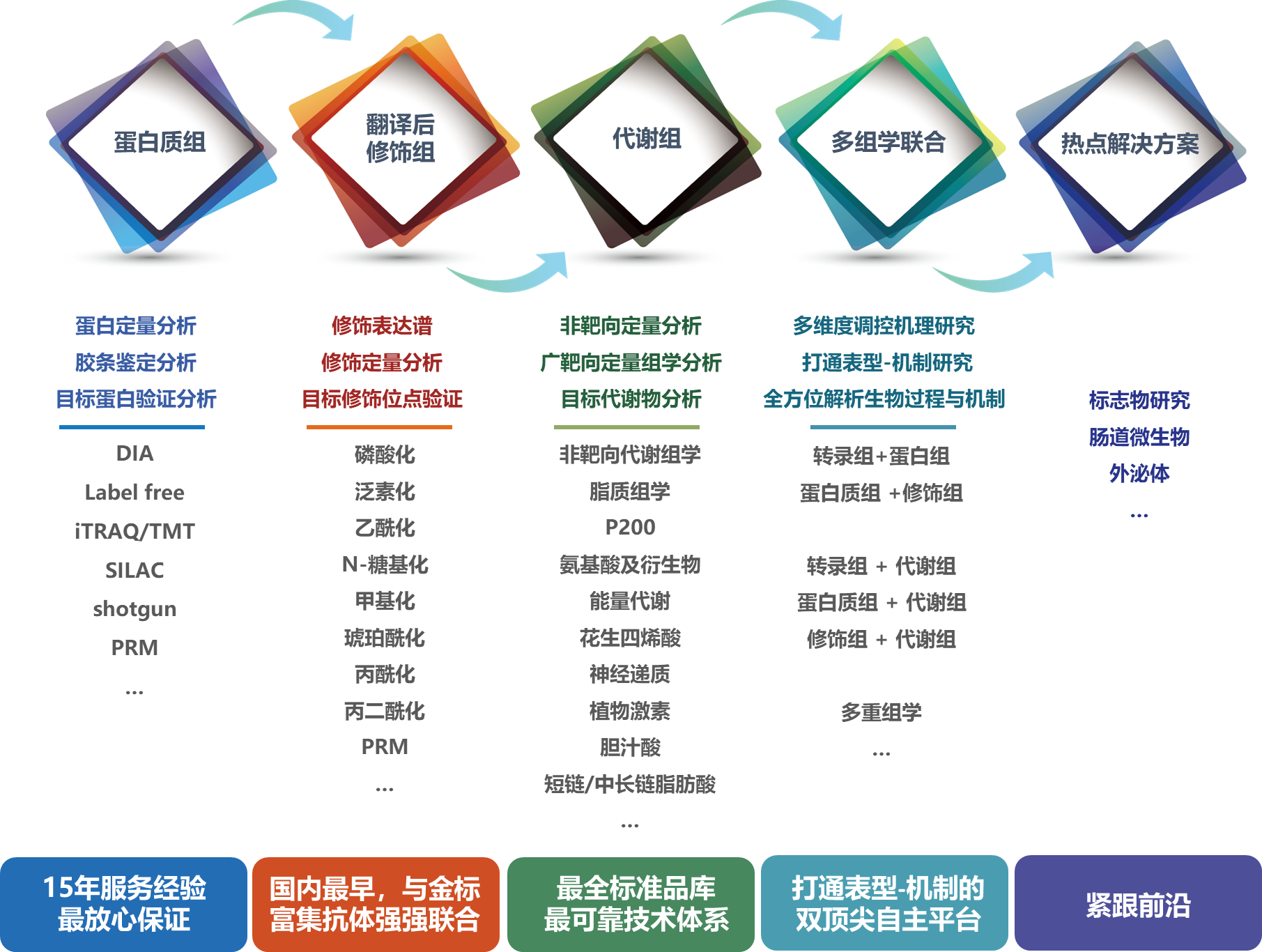
* **第三档期刊（影响因子4分以下）**

1. 组学期刊：BMC Genomics， Proteomics，Journal of Proteomics，Journal of Proteome Research，Metabolomics等
2. 综合期刊：Plos One等
3. 农林领域专业期刊: BMC Plant Biology，Journal of Agricultural and Food Chemistry，Frontiers in Plant Science等

**组学思路**：这类文章一般是最常见也是最简单的组学研究思路。1. 蛋白质组学方向：组学+差异蛋白WB或PRM验证。2.蛋白质翻译后修饰组学方向：以数据分析为主。3. 代谢组学方向：以数据分析为主。

**质谱科技服务产品概览**

中科新生命（APT）作为国内提供质谱系统解决方案的领航者，为客户提供蛋白质组、翻译后修饰组、代谢组等一系列解决方案。目前已与国内500多家科研院校、400多家医院开展过合作，年分析样本量上万例。



代表性客户文献：

[1] A paralogous decoy protects *Phytophthora sojae* apoplastic effector PsXEG1 from a host inhibitor. **Science** 蛋白质组学

[2] Lysozyme-Assisted Photothermal Eradication of Methicillin-Resistant Staphylococcus aureus Infection and Accelerated Tissue Repair with Natural Melanosome Nanostructures. **ACS Nano** 蛋白质组学+ PRM验证

[3] Natural alleles of a proteasome α2 subunit gene contribute to thermos tolerance and adaptation of African rice. **Nat Genet** 修饰组学

[4] Protein phosphatase 2A has an essential role in promoting thymocyte survival during selection. **Proc Natl Acad Sci U S A**修饰组学

[5]One-Carbon Metabolism Supports SAdenosylmethionine and Histone Methylation to Drive Inflammatory Macrophages. **Molecular Cell** 代谢组学

[6] Maize Oxalyl-CoA Decarboxylase1 Degrades Oxalate and Affects the Seed Metabolome and Nutritional Quality. **Plant Cell** 代谢组学

[7] Fecal Microbiota Transplantation Beneficially Regulates Intestinal Mucosal Autophagy and Alleviates Gut Barrier Injury. **mSystems** 肠道微生物