一. 数据下载问题

1.网页报告乱码或不可见？

在我们提供给老师的结果中包括三部分：网页报告（Report 压缩包）、完整的分析结果（summary）和原始数据（Data）；在解压 Report 压缩包时需要将 src 文件和 html 文件一起解压，移动到其他地方时也要同时移动，这样才能正常显示。单独解压 html 或者只移动 html 均会使报告打开显示不正常。

2.为什么我的报告中的有些分析结果不完整？

报告中仅展示部分结果，例如，差异分析结果仅展示属水平其中一个比较组的差异分析 结果；表格仅展示前 100 行的信息。完整的分析结果在 summary 文件中进行展示，全款到账后会释放给您。

3.MD5 文件是什么？

MD5 文件是用来核对下载数据是否完整的。您下载完数据后，不要做任何数据改动，按照链接（https://www.omicstudio.cn/doc/1142）进行 MD5 值校验。若生成的 MD5 值与下载的 MD5 值不一致，说明下载的数据与我们释放的数据有差别，需要重新下载，确保下载数据完整无缺。

4.云平台数据没有了，能重新发送给我吗？

自结题之日起，数据仅在云平台免费保存 3 个月，3 个月后会自动删除，您需要在收到结题邮件后及时前往云平台下载数据并做好备份（尤其是原始数据）。如果没有下载需要恢复数据的，收费标准为 3000 元/T/项目，不满 1T 按照 1T 计算，详情可直接与对应销售或项目管理联系。

二．数据常见疑问整理

1. Report、Summary、Data 分别是什么?

在项目完整结题后，老师一共会收到 Report、Summary 和 Data 这三部分数据。其中，Report 中主要是网页版的报告，里面有一个 summary_part 是辅助网页版报告展示结果的部分数据。Summary 是完整的分析后数据，可以配合网页版报告对应查看 Summary 的结果，Data 是原始的下机数据，**在发表文章时需要将这部分数据上传至公共数据库**，请老师及时下载保存。上述三部分结果下载完成后，**可以进行 md5 检验，保证下载结果的完整**。

2. QC 是什么，有什么作用?

在提取样本中的代谢物后，从每个样本取等量代谢物提取液混合并均分成各 QC 样品，QC 样本是技术重复。因为代谢组检测是各个样本同一批次先后上机，为了判断在检测过程中仪器是否一直稳定，将制备的 QC 样本穿插上机。所以，QC 样本主要用于质控，为了检测仪器的稳定性。

3. 代谢离子、代谢物的区别是什么?

在质谱仪中检测的物质的代谢离子，数据分析时对代谢离子进行注释，注释的结果是代谢物。使用质荷比进行注释的结果是一级代谢物，使用峰谱图进行注释的结果是二级代谢物，所以二级代谢物的结果比一级代谢物的结果更准确，可以优先选择二级代谢物进行分析。

4. pos 和 neg 是什么意思?

pos 和 neg 表示的是两种扫描模式。离子化之后的代谢物有些偏向带正电，有些偏向带负电，为检测到更多的代谢物，我们会分别进行正离子模式和负离子模式两种模式的扫描。两种扫描模式数据均可使用，在分析时已合并。

5. 代谢离子的 ID 有什么意义嘛，可以通过这个 ID 去哪个数据库中查找代谢物信息吗?

代谢离子的 ID 是根据扫描模式、质荷比（m/z）和保留时间（RT）进行命名的。这个离子 ID 是对代谢离子的编码，没有实际意义，也不能在公共数据库中查找，如果需要查找，可以直接在公共数据库中查找代谢离子对应的代谢物的名称或者代谢物在数据库中 ID。

6. 代谢物注释使用的数据库是什么？该怎么参考?

代谢物鉴定使用代谢物标准品二级质谱图谱库，再结合 Massbank、HMDB、LipidMaps、Lipidblast 等公共数据库整理而成的 in-house 二级谱图库进行，并也会提供 HMDB 一级谱图库鉴定信息。由于一级鉴定结果准确度较低，**建议参考 idms2 匹配结果。**

7. 试验样本不是人类样本，用 HMDB 数据库是否可信?

代谢物是不区分物种的，所以 HMDB 数据库是可以使用的，HMDB 数据库收录代谢物信息非常丰富，是最常用的代谢组数据库之一，有些客户是植物样本也是可以用 HMDB 数据库鉴定结果，也是可以正常发表文章的。

8. HMDB 鉴定结果中的 Superclass 和 Class 是什么?

Superclass 和 Class 都是 HMDB 数据库中的分类。在 HMDB 数据库中对代谢物进行分类，其中，一级分类是 Kingdom，二级分类是 Superclass，三级分类是 Class，四级分类是 Subclass，这四级逐级细化，逐步精确代谢物的分类。例如，HMDB0000032 这个代谢物，

对应的一级分类是有机物，二级分类是脂质和类脂质类物质，三级分类是类固醇和类固醇衍生物，四级分类是胆甾烷类固醇。

Chemical Taxonomy	
Description	Belongs to the class of organic compounds known as cholesterols
Kingdom	Organic compounds
Super Class	Lipids and lipid-like molecules
Class	Steroids and steroid derivatives
Sub Class	Cholestane steroids
Direct Parent	Cholesterols and derivatives

9. KEGG 结果中的 level1、level2 是什么?

Level1、level2 都是 KEGG 数据库中对代谢通路的分类。在 KEGG 中，将代谢通路划分为 7 类，分别是：新陈代谢（Metabolism）、遗传信息处理（Genetic Information Processing）、环境信息处理（Environmental Information Processing）、细胞过程（Cellular Processes）、生物体系统（Organismal Systems）、人类疾病（Human Diseases）和药物研发（Drug Development），其中每类代谢通路又系统分为二、三层。例如，Metabolism 划分成的氨基酸代谢，脂质代谢等是 level2，第三层（level3）即为每条具体通路。

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for:

1. **Metabolism**
   Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**
7. **Drug Development**

KEGG PATHWAY is the reference database for pathway mapping in **KEGG Mapper**.

10. 定量结果的表格中，前面是离子 ID，怎么查找对应代谢物名称?

在所有含有表达量数据的表格中，找到表头中写有“metabolite 或 ms2.name”字样的一列，即为该代谢离子对应的代谢物名称，后面对应的各个样本的数字就是该代谢物的表达量。

	B	C	D	E	F	G	H	I	J
		RT	IsAnnotat	IsMS2ident	MS2Metabolite	MS2superclass	MS2class	MS2hmdb	MS2kegg
	). 01207	9. 799883	0	0	–	Unknown	–	–	–
	). 01196	9. 318383	0	0	–	Unknown	–	–	–
	). 01249	0. 65788	0	0	–	Unknown	–	–	–
	'0. 0124	1. 170371	0	0	–	Unknown	–	–	–
	). 01235	2. 989067	0	0	–	Unknown	–	–	–
	). 06499	0. 857058	1	1	1-Pyrroline	Organoheterocyclic co	Pyrroline	HMDB0012	C15668
	). 10552	5. 033883	0	0	–	Unknown	–	–	–
	). 51417	0. 650285	0	0	–	Unknown	–	–	–

11. 结果里面代谢物名称为 A1,但是文献里面的为 A2，但是在 kegg 数据库中的 id 号是同一个，这个是为什么?

这个需要具体情况具体分析。一个代谢物可能有多个名称，例如 3-磷酸葡萄糖和葡萄糖-3-磷酸这是一个物质，只是说法不一致；如果是 FA 16:1 和 FA 17:1 这种，二者对应的 KEGG ID 都是 C00162，但是这是两种物质。KEGG ID 相同说明这些代谢物在 KEGG 数据库中行使同一功能，并不代表代谢物是一个。

12. 代谢物的名称可以在哪些数据库中查找到？

代谢物名称可以在 Pubchem(https://pubchem.ncbi.nlm.nih.gov/), HMDB(https://hmdb.ca/), KEGG(https://www.genome.jp/kegg/)等数据库中进行查找，可以直接输入代谢物名称或者代谢物在对应数据库中的 ID，即可搜索出相关信息。

13. 代谢物名称出现中文字符，为什么？

代谢物名称中出现中文字符是因为数据输出到 Excel 时出现了转码错误，可以将带有中文字符的代谢物的名称复制到 Pubchem(https://pubchem.ncbi.nlm.nih.gov/)中查找真实的名称。

14. 代谢物的表达量有单位吗？

非靶向代谢组检测的代谢物的表达量是没有单位的，这个定量是相对定量的结果，非靶向代谢组采用该代谢物的峰面积作为其表达量。

15. P 值和 q 值什么区别？

P 值是统计学差异显著性检验指标，一般 p<0.05 认为差异显著。q 值是校正后的 P 值，是对 P 值进行了多重假设检验，能更好地控制假阳性率，我们使用 BH 算法进行 P 值的矫正。P 值和 q 值在分析结果中均有提供，q 值相对来说会更加严格，一般使用 P 值即可。

16. VIP 值是什么？

VIP（Variable Important for the Projection））是 PLS-DA 模型的变量投影重要度，描述的是每个变量（每个代谢离子）的差异对模型的贡献度，一般这个值越大证明差异越大，拆分析一般阈值为 VIP＞1。

17. PCA、PLSDA、OPLSDA 区别？

PCA 是做常见的降维方法，是一种无监督的模式。PLSDA 是有监督的模式，利用偏最小二乘回归构建模型，对数据降维。OPLSDA 在 PLSDA 基础上增加正交分析，在代谢组分析中 PLSDA、OPLSDA 均经常使用。

18. KEGG 富集分析中 P 值怎样计算的？

KEGG 通路的 P 值主要根据四个数值进行**超几何检验**计算而来，具体的算法参考下方公式：

$$P = 1 - \sum_{i=0}^{S-1} \frac{\binom{B}{i}\binom{TB-B}{TS-i}}{\binom{TB}{TS}}$$

S：注释为特定 KEGG 通路中的差异代谢物个数

TS：有 KEGG 通路注释的差异代谢物总数

B：注释为特定 KEGG 通路中的代谢物个数

TB：有 KEGG 通路注释的代谢物总数

P value：超几何检验的 p 值

19. KEGG 富集分析中 P 值表示什么？P 大于 0.05 是不是不能用？

KEGG 通路 P 值表示的是这条通路的显著性，P 值主要是由这条通路上鉴定到的差异代谢物和这条通路上总代谢物个数决定的。

筛选通路时不用只考虑 P<0.05 的通路，因为所有通路都是差异代谢物富集的通路，在所有富集到的结果中筛选出关注的通路即可进行后续分析。

20. 样本是植物样本，富集分析结果为什么出现了动物的相关通路？

我们在对差异代谢物进行富集分析时使用的是 KEGG 总库，不区分动物植物，所以有些植物样本的结果里面可能有动物的结果。这时您可以在富集分析结果中剔除动物的部分，再联系售后同事重新出图。

21. 比较组填反了，对后面结果有哪些影响？

比较组的顺序反了，只涉及到个别结果的颠倒，例如，差异代谢物的上下调统计、火山图，热图、PCA、PLSDA、富集分析等不受影响。所以，如果在拿到结果后发现比较组设置颠倒了，可以联系售后同事调整，不用重新分析。

22. 代谢组可以做多组比较吗？与两两比较有什么不同？

代谢组可以进行多组比较，多组比较的阈值标准是:Anova p<0.05,VIP≥1。

两两比较主要关注处理组与对照组相比哪些代谢物发生了变化，多组比较主要关注多组之间代谢物的变化趋势，二者侧重点不同，可以根据试验设计进行选择。

23. 非靶向代谢和靶向代谢有什么区别？

非靶向代谢组在于尽可能多的检测代谢物，靶向代谢组在于只检测关注的一类代谢物的含量。一般用非靶作为代谢物的初筛，然后进行靶向进行验证和绝对定量。在进行靶向的实验之前，需要联系销售同事进行评估。

从数据角度，非靶向代谢组检测得到的代谢物的表达量是相对定量结果，靶向代谢物检测得到的代谢物的表达量是绝对定量的结果。

24. 脂质和常规非靶有什么区别？

常规非靶理论上可广谱检测样本中包含脂质和非脂质的所有代谢，但主要集中在水溶性代谢物、部分极性较强的脂溶性代谢物。非靶脂质的重点在于检测脂溶性代谢物。如果没有方向，想广谱看看可以检测到哪些代谢物，或者组间可以检测到哪些差异代谢物，建议做常规非靶。如果重点关注脂质类物质的差异，建议做脂质组。

25. 脂质组中 PC16:0 和 PC18:0 是同一个物质吗？

PC16:0 和 PC18:0 不是同一个物质，16:0 表示的是有一个支链，支链上有 16 个 C，其中没有不饱和键，16 对应的是该 C 链上 C 原子的个数，0 对应的是 C 链上的不饱和键个数。所以 16:0 和 18:0 显然不是同一个物质。而且 PC16:0 表示的是一类代谢物，因为在非靶水平上只能检测出 C 链上的 C 原子和不饱和键个数，无法检测到出具体的排列方式。

26. 脂质组鉴定的物质分类 PC、PG 等分别表示什么？

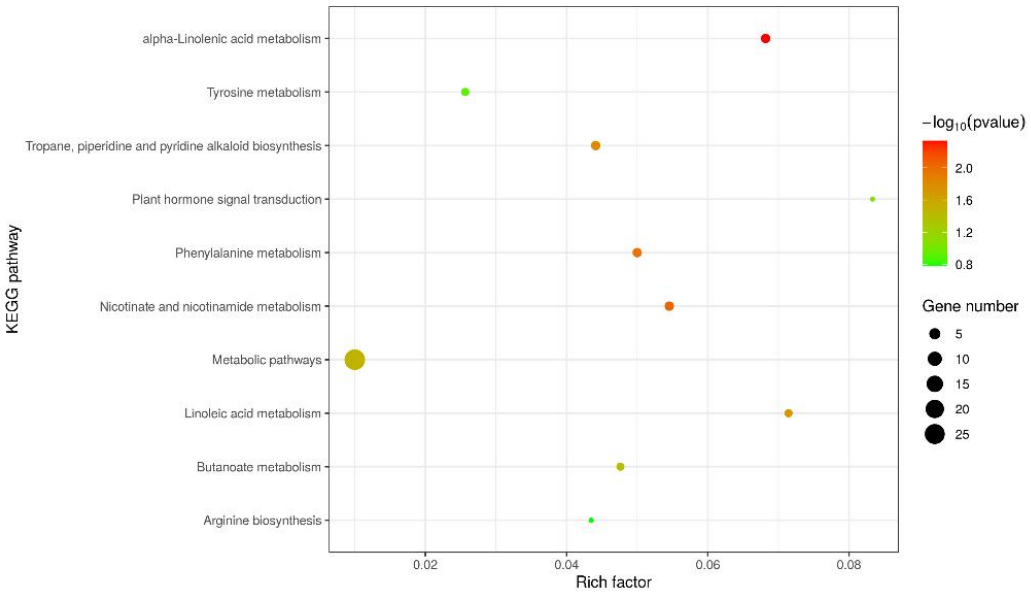
在非靶脂质的结果中，代谢物的分类通常是 PC、PG 等，其中 PC 表示的是磷脂酰胆碱，PG 表示的是磷脂酰甘油，具体的分类详见下表：

Class II	物质二级分类	Class II	物质二级分类
BA	胆汁酸	LPE	溶血磷脂酰乙醇胺
BMP	双单酰基甘油磷酸酯	LPE-P	溶血磷脂酰乙醇胺(烯醚键)
CAR	酰基肉碱	LPG	溶血磷脂酰甘油
CASE	谷甾醇酯	LPI	溶血磷脂酰肌醇
CE	胆固醇脂	LPS	溶血磷脂酰丝氨酸
Cer	神经酰胺	MG	单甘油酯
CerP	1-磷酸神经酰胺	MGDG	单糖甘油二酯
Cert	植物神经酰胺	PA	磷脂酸
Cho	胆固醇	PC	磷脂酰胆碱
Class II	物质二级分类	PC-O	磷脂酰胆碱(醚键)
CoQ	辅酶Q	PE	磷脂酰乙醇胺
DG	甘油二酯	PE-O	磷脂酰乙醇胺(醚键)
DGDG	二糖甘油二酯	PE-P	磷脂酰乙醇胺(烯醚键)
DG-O	甘油二酯(醚键)	PG	磷脂酰甘油
Eicosanoi	氧化脂质	PI	磷脂酰肌醇
FFA	游离脂肪酸	PMeOH	磷脂酰甲醇
HexCer	糖鞘脂	PS	磷脂酰丝氨酸
LNAPE	N-酰基-溶血磷脂酰乙醇胺	SM	鞘磷脂
LPA	溶血磷脂酸	SPH	鞘氨醇
LPC	溶血磷脂酰胆碱	TG	甘油三酯
LPC-O	溶血磷脂酰胆碱(醚键)	TG-O	甘油三酯(醚键)

27. 一级二级代谢物区别（MS1 与 MS2 的区别）

一级代谢物：利用开源软件 metaX 对物质的质荷比(m/z)与 KEGG 进行匹配得到一级代谢物鉴定结果（由于存在不同物质具有相近质荷比的情况，鉴定结果会有一个 m/z 对应多个代谢物的现象）。二级代谢物：将不同物质在质谱仪里面进行碎裂，利用碎片信息生成二级谱图，与公共数据库的标准品二级谱图和我们的 in-house 图谱库进行匹配并对匹配结果评分，最终得到代谢物的二级鉴定结果，二级代谢物鉴定更加准确。

28.生信分析 KEGG 气泡图中富集因子的含义？该如何选择受到显著影响的通路进行研究？



KEGG 通路富集分析

横坐标表示每条 KEGG 通路的富集因子，富集因子（rich factor）指生信分析文件夹中 kegg 表格的 count/pop hit，即参与某 KEGG 通路的差异代谢产物的数目占该通路注释到的代谢产物的比例，一般情况下，KEGG 通路富集结果中 P 值越小（P<<0.05），统计学上 KEGG 通路富集越显著，而 KEGG 通路下包含的差异表达代谢物数目在某种程度上反映实验设计中生物学处理对各个通路的影响程度大小，因此可以结合两方面因素，选择较为感兴

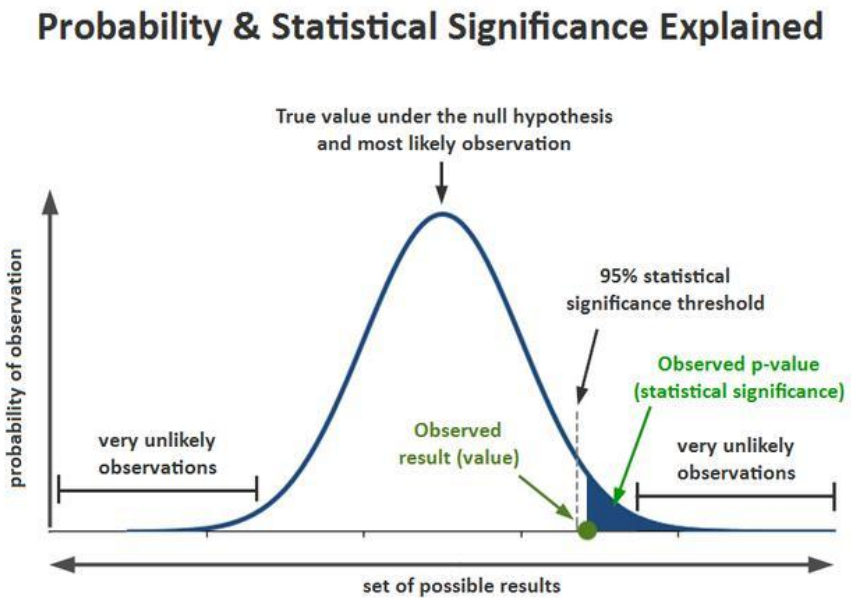
趣的代谢或信号转导途径以及显著性影响这些途径的差异表达代谢物进行后续生物学实验验证或机制研究。

29.差异代谢物有 a 个，富集到通路的差异代谢物共有 b 个（a＞b）？

鉴定到的差异代谢物有 a 个，但是并不是每个代谢物都有 KEGG 通路注释，富集分析是对有通路注释的差异代谢物进行分析,所以可以用于富集分析的差异代谢物个数要少于鉴定到的差异代谢物个数。

30.代谢组学常用的显著性检验方法：

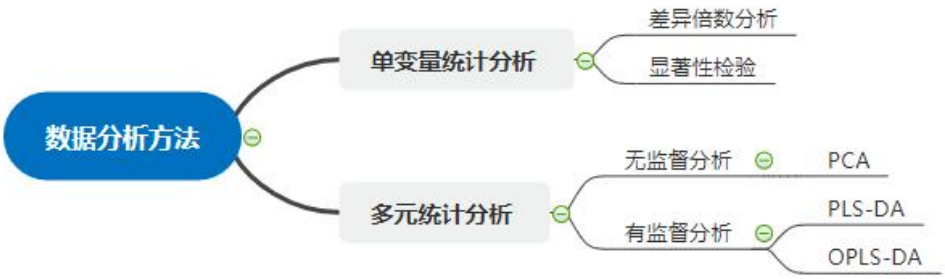
p 值是一个概率，反映某一事件发生的可能性大小，用于区分该变量是否具有统计显著性，通常认为 p<0.05 具有统计学意义。常用的检验方法有 t-test、方差分析（Analysis of Variance， ANOVA）。t 检验一般适用于两组差异比较，在多组的情况下就要用到 ANOVA 方差分析。



31.单变量分析方法-差异倍数分析在代谢组学两两比较中是较为常见的，但多组比较为什么没有呢？

差异倍数（Fold Change，简称 FC 值）分析即根据代谢物的相对定量或绝对定量结果，计算某个代谢物在两组间表达量的差异。差异倍数作为上下调的一个标准，假设比较组为 AvsB，计算方式为：FC=B/A，FC 大于 1 为上调，小于 1 为下调（这个标准不是固定的，也可以设置的更为严格一点，比如调整为 1.2 倍、1.5 倍或者 2 倍，这三种阈值在代谢组研究相关文章中是较为常见的）。我们说上下调，一般都是指和某一组相比，另一组上调或者下调，三组或者多组的时候是无法定义和哪组相比其他几组高或者低的，因此差异倍数是在两两比较中产生的。

32.代谢组学常用到的差异代谢产物的数据分析方法：



单变量分析方法是简单常用的实验数据分析方法。在进行两组样本间的差异代谢物分析时，常用的单变量分析方法包括差异倍数分析（Fold Change Analysis，FC Analysis）、T 检验，以及综合前两种分析方法的火山图（Volcano Plot）。

多元统计分析中无监督分析有主成分分析（PCA），而有监督分析方法主要是偏最小二乘判别分析（PLS-DA）和正交偏最小二乘判别分析（OPLS-DA）。

VIP（Variable important in projection）是(O)PLS-DA 模型变量的变量权重值，来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代谢物。

由于代谢组数据具有多维且某些变量间高度相关的特点，运用传统的单变量分析无法快速、充分、准确地挖掘数据内潜在的信息，因此一般采用多元统计分析方法，可以在较大程度保留原始信息的基础上将高维复杂的数据进行“简化和降维”，建立可靠的数学模型对研究对象的代谢谱特点进行归纳和总结。

因此代谢组学推荐使用单维和多维的方法进行结合，有助于我们从不同角度观察数据，得出结论。所以选择 P 值小于 0.05 与 VIP 值大于 1 作为常见的差异代谢物筛选标准。

33. 代谢组学中 LC-MS 与 GC-MS 数据的区别：

TYPE	LC-MS	GC-MS
数据库	METLIN、MassBank、mzCloud与自建库	NIST、Fiehn、GMD
离子源	**ESI**、APCI ……	**EI**、**CI**、NCI……
数据形式	正离子模式+负离子模式	单一离子模式（正/负）
分析物质	适用范围广	易挥发，热稳定性物质

1）LC-MS 根据电离方式不同，可分为电喷雾离子源（ESI）和大气压化学电离源（APCI） 2 种工作方式；GC-MS 有电子轰击电离（EI）、正化学电离（CI）、负化学电离（NCI）3 种电离方法，其中前两者较常用。

2）LC-MS 是在正、负离子两种模式下工作的，得到的数据形式也是不一样的，而对代谢物的统计学分析时也是分开的，但在代谢通路分析时（或者合并分析时），会将正负离子结合，有重复时选择两种模式中响应较高的一个模式。

3）GC-MS 通常只能在单一离子模式下工作，得到的数据模式非负即正，可根据实际的离子源进行判断，因此在分析时工作量就少了一半。再加上由于扫描离子范围的差别，LC-MS 获得的数据量明显更多。

相比于 GC-MS，LC-MS 一般无需衍生处理，分析平行性更好，更适合大规模样本的分析。

34. 血液样本做代谢组学分析，血清样本和血浆样本哪一个比较好？

血清血浆都是血液样本处理后得到的样品,现有文献报道血清血浆中代谢物种类及丰度确实不同，但对于研究而言，并没有明确表明哪种样本类型优于另一种，所以在选择血清或者血浆时，只要在收样时保证统一即可，且血液样本最好是选择 EDTA 或肝素抗凝的血浆比较好。收集过程需要避免溶血，样收集后应保存在－80℃条件下，并且避免反复冻融。

35. 靶向代谢是如何进行定性和定量的？

靶向定性是根据代谢物的母离子和子离子分子量，通过质谱 MRM 模式进行定性。靶向绝对定量是根据代谢物的实际检测峰面积与标准品的峰面积进行换算得到的。

36. 脂质组学的命名规则问题？

脂质的命名中，数字代表碳长度及双键个数，例如 FA(3:0/20:2)，表示有一个长度为 3 和两个长度为 20 的碳链。长度为 20 的碳链中有两个不饱和键。

三、数据挖掘思路

1. 怎样快速筛查数据？

查找重点在拿到代谢物检测的数据后，首先查看 QC，因为 QC 是真正意义上的技术重复，如果 QC 都是稳定结果较好即可排除仪器的影响。然后直接看差异代谢物的结果，查看在比较组中有哪些差异代谢物，在这种处理下主要是哪几类代谢物发生变化，这些差异代谢物可能通过哪些通路发挥作用。在此次的检测中有哪些代谢物的变化与前人文献研究得到相同的结果，又有哪些新发现的代谢物可能影响这种表型的变化。

2. 拿到结果后重点看哪部分数据？

在拿到完整结果之后，可以将重点放在差异分析部分。

关于差异分析部分。我司提供多条思路：1.常规差异代谢物筛选及富集分析、网络图分析。2.代谢组 GSEA 富集分析及相关绘图展示。3.差异代谢物间相关性分析及相关性网络图等分析。

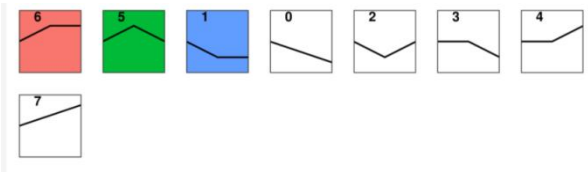
具体数据挖掘可以根据您的数据情况来有侧重选择。根据一般思路，**优先看常规差异代谢物筛选及富集分析、网络图分析部分**。如果此部分差异代谢物较少、富集到通路与研究无关等，分析结果不理想，可以**再重点关注代谢组 GSEA 富集分析部分**，GSEA 分析使用所有代谢物做富集分析,在算法上对于变化幅度不大没达到显著差异标准但生物学意义可能很重要的代谢物、通路途径富集更有优势，是常规富集分析很好的补充。差异代谢物间相关性分析一般作为辅助，可探究差异代谢物间表达量之间是否有关联，**相关性分析对应探索未知的调控关系帮助较大。**

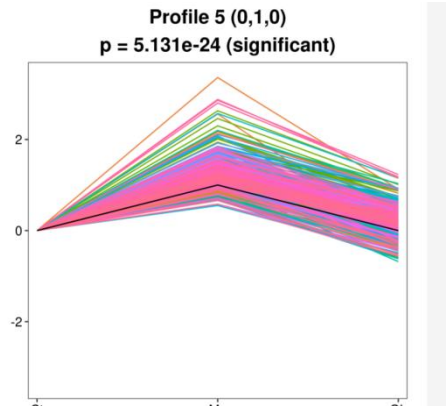
3. STEM 研究重点是根据不同的时间点或者治疗模型，筛选出关键的差异代谢物？

如果样本之间有一定的规律，例如，不同时间节点取的样本（如：0h-4h-8h 不同时期取样）、构建治疗模型按照治疗趋势取的样本（如：对照组-损伤组-保护组），这种代谢物在

组间有预期的变化趋势时，可以考虑根据多组比较结果进行趋势分析筛选数据。

将多组之间获得的差异代谢物进行趋势分析，需要整理代谢物在各个组中表达量的平均值，按照顺序排列，上传至云平台“趋势分析”这个云工具即可进行分析，根据分析结果找出预期变化趋势中的差异代谢物，再根据通路或者分类进行后续分析。



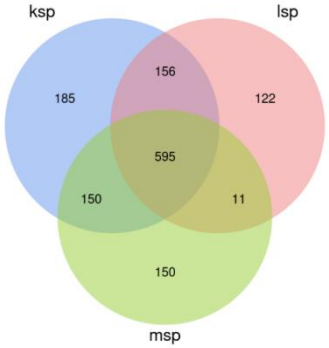


4. Venn 研究重点是找出不同处理条件下，发生显著变化的差异代谢物？

如果样本是不同的处理，想找出不同处理条件下可以发生显著性差异的代谢物，可以选择 Venn 筛选数据进行分析。

可以整理不同处理条件下发生显著变化的代谢物，根据 Venn 图筛选，找出在不同处理条件下均发生显著变化的代谢物，也可以获得只某一种处理条件下显著变化的代谢物。

还有一种情况，当趋势分析结果非常多，不便于后续分析时，可以选择 Venn 的思路筛选，因为 venn 这种思路是基于两两比较，趋势分析是基于多组比较，而两两比较的阈值比多组比较阈值更严格。



5. 结合富集分析研究重点是找出与某些通路（例如：花青素生物合成）相关的差异代谢物？

如果根据文献和科研背景，已经有预期的代谢通路，可以在分析结果中直接筛选通路，如果差异代谢物的富集分析结果中有关注的通路，直接提取通路及通路中的代谢物，对代谢物的表达量、分类、上下游基因等进行分析。如果差异代谢物富集分析结果中没有关注的通路，可以返回所有代谢物的富集分析结果中查看，如果所有代谢物的富集分析结果中有关注的通路，提取通路中的代谢物，查看表达量信息并适当调整阈值；如果所有代谢物的富集分

析结果中没有关注的通路，则检测出的代谢物并不能富集到预期的通路中，可以再查阅文献从其他角度入手分析数据。

6. 想要筛选出生物标志物？

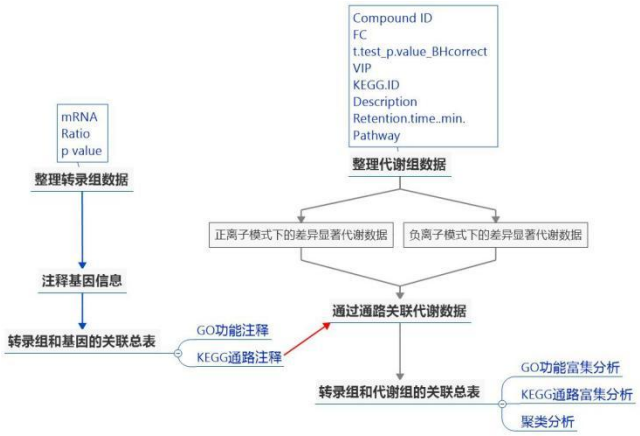
一般筛选生物标志物，可以结合文献直接筛选在一个组中代谢物表达量显著高于其他组的代谢物，也可以使用 ROC 曲线评估生物标志物的工作效率，还可以根据随机森林或 WGCNA 等挖掘变量之间复杂的非线性的关系。通常筛选生物标志物通常是临床样本，对样本数量有一定要求。ROC 要求每组样本数量大于 10 例，随机森林要求每组样本数量一般大于 50 例。一般认为，AUC>0.8 才是比较好的标志物。

7. 与转录组如何关联？

通常，转录和代谢联合分析是通过 KEGG 代谢通路将两组学数据联合起来，找到同一生物进程（KEGG Pathway）中发生显著性变化的基因和代谢物，快速锁定关键基因和代谢物。

如果基于通路不能获得有用的信息，也可以使用相关性进行关联，可以筛选出关注的差异基因和差异代谢物，计算二者之间的相关性，分析哪些基因和代谢物有同步的变化趋势。

对于分析结果的可视化，可以选择网络图，相关性聚类热图，相关性网络图等等。

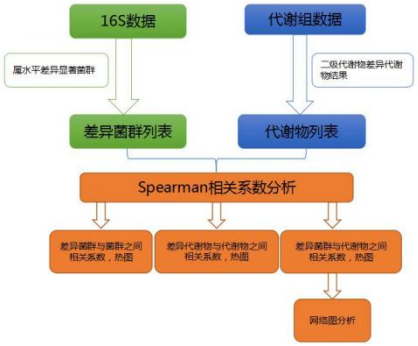


8. 与蛋白组如何关联？

蛋白与代谢的联合分析思路同转录代谢联合分析，都是基于通路或者相关性，只是找出差异蛋白和差异代谢物的关系。

9. 与 16S 如何关联？

16S 与代谢的联合分析主要是基于 Spearman 相关性，获得差异菌群与差异菌群，差异代谢物与差异代谢物，差异代谢物与差异菌群之间的关系。基于计算结果选择合适的筛选条件，获得最终的差异代谢物与差异代谢物的相关关系及网络图等。



10. 与其他项目的联合分析?

除上述常见的联合分析之外，还有 m6A 与代谢，miRNA 与代谢，ceRNA 与代谢等，这些项目与代谢联合分析整体思路可以根据 KEGG 代谢通路进行关联，以 miRNA 与代谢为例，找出 miRNA 靶向的基因参与的代谢通路，再与代谢组数据关联。

11. 后续试验验证

后续的试验验证，可以选择验证代谢物或者上游基因或酶。

验证代谢物可以选择靶向代谢组，或者买相应的试剂盒检测代谢物。在非靶的基础上初筛差异代谢物，筛选出一类关注的代谢物，进行靶向代谢组的检测。值得注意的是，筛选出的代谢物是否可以进行靶向检测，需要先联系销售经理评估。也可以进行一些体外细胞培养或者动物模型的验证。

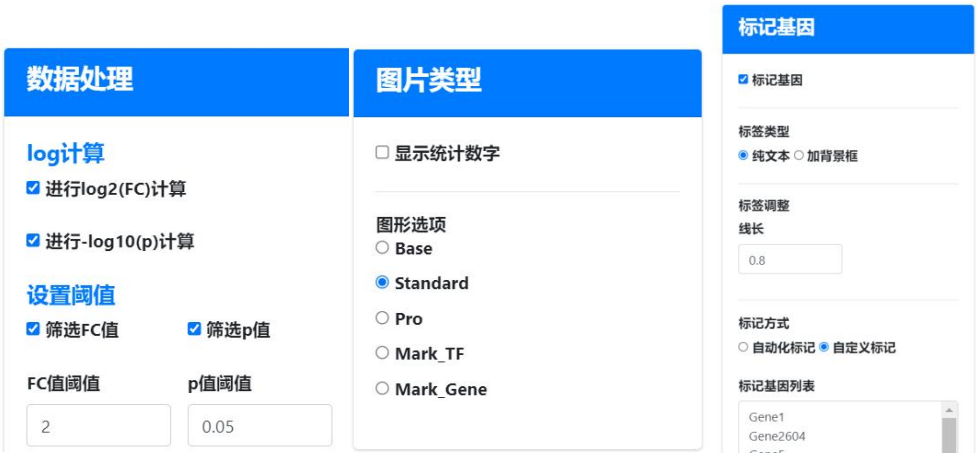
验证上游基因或酶，可以根据代谢通路寻找其上游的基因，通过过表达或者敲除试验，观察表型。

四、关于作图

代谢组报告中的图片基本都可以通过联川生物云平台（https://www.omicstudio.cn/index）进行绘制，如果需要重新绘制，可以在分析结果中整理出对应的数据，上传至云平台绘制。

1. 火山图用什么数据绘制的?

火山图需要使用的数据是 MS2Metabolite（代谢物名称），ratio（差异倍数），p 值或者 q 值（具体使用哪个数据请参考报告中关于差异筛选的描述），将这三列整理在一个 excel 中，在云平台选择“高级火山图”这个云工具上传表格，数据处理选择如下，阈值根据报告中设置，图片类型可以选择“Standard”，可以选择标记代谢物，如果标记代谢物可以在“标记基因”这个模块上传需要标记的代谢物名称，后面的颜色字体等可以根据自己的喜好进行调整，待所有调整结束后，下载保存图片即可。

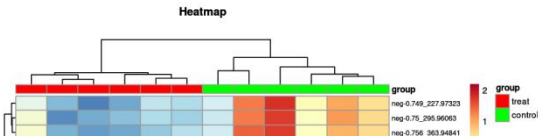


2. 热图用什么数据绘制的？

热图需要使用的数据是 MS2Metabolite（代谢物名称）和该代谢物在各个样本中的表达量数据。在云平台选择“高级热图”这个云工具上传表格，数据处理这里选择 log10 处理和 Z-score 归一化，其他的均可以按照自己的喜好添加修改，包括是否显示行列名，是否按行按列聚类，是否对样本进行分类，是否按照代谢物的分类进行分类等。

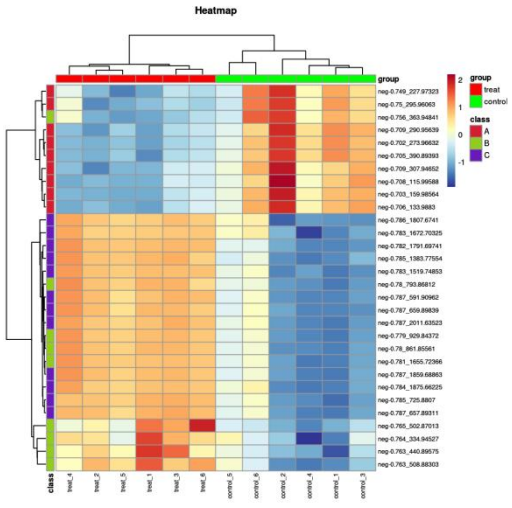
对样本进行分类，需要整理一张表格，注明每个样本属于哪个分组，上传列注释表格即可实现。

ID	group
treat_1	treat
treat_2	treat
treat_3	treat
treat_4	treat
treat_5	treat
treat_6	treat
control_1	control
control_2	control
control_3	control
control_4	control
control_5	control
control_6	control



按照代谢物的种类进行分类，表格同列注释的表格，第一列是代谢物的名称，第二列是代谢物的分类，上传至行注释即可实现。待所有调整结束后，下载保存图片即可。

ID	class
neg-0.702_273.96632	A
neg-0.703_159.98564	A
neg-0.705_390.89393	A
neg-0.706_133.9883	A
neg-0.708_115.99588	A
neg-0.709_290.95639	A
neg-0.709_307.94652	A
neg-0.749_227.97323	A
neg-0.75_295.96063	A
neg-0.756_363.94841	B
neg-0.763_440.89575	B
neg-0.763_508.88303	B
neg-0.764_334.94527	B
neg-0.765_502.87013	B
neg-0.779_929.84372	B
neg-0.78_793.86812	B
neg-0.78_861.85561	B
neg-0.781_1655.7236	B
neg-0.782_1791.6974	C
neg-0.783_1519.7485	C
neg-0.783_1672.7032	C



3. PCA 图用什么数据绘制的？

PCA 需要两张表格，一张是所有的代谢离子及其表达量数据，注意这张表格上不能有表达量为空的代谢离子，如果有一行都是空值的代谢离子，将这一行删除即可，否则分析时会报错；另一张是各个样本的分组文件，如下：

ID	treat_1	treat_2	treat_3	treat_4	treat_5	treat_6	control_1	control_2	control_3	control_4	control_5	control_6
pos-5.196_	7.998E+09	7.653E+09	8.099E+09	5.573E+09	5.387E+09	7.13E+09	4.6E+09	5.191E+09	6.783E+09	4.61E+09	6.854E+09	6703058433
pos-10.05	3.989E+09	2.502E+09	3.069E+09	2.207E+09	2.351E+09	2.913E+09	2.39E+09	1.961E+09	2.348E+09	1.809E+09	2.881E+09	2505573020
pos-9.234_	1.552E+09	1.461E+09	1.097E+09	2.474E+09	2.866E+09	1.412E+09	1.693E+09	1.101E+09	1.034E+09	1.881E+09	1.395E+09	1034367829
pos-9.234_	1.457E+09	1.372E+09	1.03E+09	2.313E+09	2.67E+09	1.325E+09	1.584E+09	1.034E+09	974280113	1.753E+09	1.31E+09	976504553.8
pos-8.361_	1.201E+09	1.138E+09	936344025	1.926E+09	2.084E+09	1.177E+09	1.271E+09	940376877	983252995	1.475E+09	1.034E+09	800896777.4
pos-3.516_	993092023	1.286E+09	761975571	1.988E+09	1.636E+09	1.058E+09	1.216E+09	746805870	969202246	1.11E+09	936294325	668356800
pos-9.938_	2.57E+09	2.418E+09	2.514E+09	1.773E+09	1.69E+09	2.613E+09	1.042E+09	1.874E+09	1.971E+09	1.257E+09	2.19E+09	1800933798
pos-10.08	2.342E+09	1.604E+09	1.579E+09	1.063E+09	1.288E+09	1.968E+09	699682554	1.51E+09	2.033E+09	1.025E+09	1.49E+09	1312177585
pos-7.898_	235465872	381421920	518938590	239992769	322676899	351574345	144300401	356307668	1.453E+09	1.159E+09	155769884	260661590.3
pos-5.347_	1.097E+09	976226061	1.337E+09	623797021	767698551	1.013E+09	987190270	602967909	706334848	900166531	1.039E+09	1268660972
pos-8.36_	719732973	667126437	580925603	1.175E+09	1.269E+09	741891299	735628813	641146078	610787302	919641455	692303030	508267475.6
pos-9.932_	1.596E+09	1.64E+09	1.617E+09	1.127E+09	1.099E+09	1.536E+09	631606399	1.226E+09	1.415E+09	770238580	1.365E+09	1193514680
pos-9.936_	1.02E+09	1.169E+09	1.216E+09	651352026	605788133	994197637	753234114	702255182	678553491	410594979	1.105E+09	669498252.2
pos-4.382_	459672569	501766553	782169000	854926072	1.146E+09	481270078	497193804	293116826	351105143	658487107	414316764	380409838.4

ID	group
treat_1	treat
treat_2	treat
treat_3	treat
treat_4	treat
treat_5	treat
treat_6	treat
control_1	control
control_2	control
control_3	control
control_4	control
control_5	control
control_6	control

在云平台选择“二维 PCA 图”这个云工具依次上传两张表格，即可出图，其他的参数可以根据自己的审美自行调整，待所有调整结束后，下载保存图片即可。

PCA 和热图都是根据表达量分析，但是 PCA 是展示样本与样本之间的关系，所以这里需要整理的是在一个样本中所有检测到的代谢离子的表达量数据，而热图可以选择关注的代谢物进行分析，可以选择部分数据。

4. 富集分析的结果太多了，后面可以怎么调整？

在差异代谢物富集分析结果中，是将所有的差异代谢物进行富集分析，所以富集出的结果比较多，如果文章中直接放这个结果可能图片较长，这时老师可以在图片对应的表格中筛选出重点关注的代谢通路或直接筛选 p 值最小的 top20，重新整理文件上传至云平台“KEGG 富集分析因子图”这个云工具即可实现。

分析，可以联系售后人员。

5. 文章中可以放置哪些图片？

代谢组的文章可以按照鉴定到的代谢物、差异代谢物、重点关注差异代谢物这个思路进行分析。对于代谢组的质控结果，一般文章中不作详细说明。鉴定到的代谢物一般只统计个数。差异分析这里可以选择差异代谢物热图、PCA 和 PLSDA 任选其一、火山图、富集分析等。热图可以选择整个比较组的差异代谢物热图，也可以选择部分关注代谢物绘制热图。PCA 和 PLSDA 可以任选其一放置，只是 PLSDA 和置换检验结果需要同时出现，若是置换检验结果证明 PLSDA 模型不成立，则不可以放置 PLSDA 结果。富集分析可以选择整体代谢物富集分析结果，也可以筛选部分通路展示。

除了分析结果中的图片，也可以根据实际研究方向选择 Venn 图，相关性热图，网络图等等，如果有需要，可以在云平台尝试分析，也可以联系售后人员处理。

上述所有图片均是可以展示，而不是一定放置，具体可以根据研究方向、参考文献和实际结果进行增减。

除这些分析内容之外，如果老师想要参考文献中的分析或者图片，可以联系售后人员评估。