

1. 代谢组学主要的研究对象及研究目标有哪些？

答： 代谢组学的研究对象主要是分子量小于 1000Da 的内源性小分子，研究方向包括：全代谢组研究（对限定条件下的特定生物样品中所有代谢组分的定性和定量），靶向代谢组研究（对某个或某几个特定组分的分析），代谢轮廓分析（对少数所预设的一些代谢产物的定量分析），代谢指纹图谱分析（不分离鉴定具体单一组分，而是对样品进行快速分类），其中全代谢组和靶向代谢组研究是目前代谢组研究的主要方向。

2. 代谢组研究相对于基因组和蛋白质组研究而言有什么不同之处？

答： 主要特点包括三方面：

基因和蛋白表达的有效的微小变化会在代谢物上得到放大，从而使检测更容易；

代谢组学的技术不需建立全基因组测序及大量表达序列数据库；

因为代谢产物在各个生物体系中都是类似的，所以代谢组学研究中采用的技术更通用。

3. 代谢组的样品重复性有什么要求？

答： 代谢组学基于多元统计分析方法进行的，在样品准备上相对于转录组和蛋白质组而言需要更多的重复数据。

一般建议植物样品：最少 6 次，建议 8 次生物学重复。

模式动物及微生物样品：最少 8 次，建议 10 次生物学重复。

临床样品：30 次生物学重复以上，如组织样品不好取样，可控制在 20 次重复以上。

4. 代谢组学在系统生物学中有什么作用？地位如何？

答： 基于组学技术的系统生物学研究内容涵盖基因组学、转录组学、蛋白质组学和代谢组学。这几种组学的研究中，基因组学主要研究生物系统的基因结构组成，蛋白质组学研究有生物系统表达的蛋白质及由外部刺激引起的差异。

系统生物学中，基因组学和蛋白质组学告诉你可能发生什么，而代谢组学则告诉你已经发生了什么。

代谢物作为生物体表型的基础能够帮助我们更直观有效地了解生命现象揭示生命本质。代谢组学本质上是指从整体上研究生物体的代谢物，是对某一生物、组织或细胞中的所有低分子量代谢产物进行定性与定量分析的一门科学。

代谢组学以指标分析为基础，高通量检测为手段，通过研究生物体内代谢物的种类与数量及其变化规律来阐述机体在正常生命状态及环境变化后的代谢过程。代谢组是系统生物学的分支学科之一。

目前，其研究内容主要包括对代谢物进行定性或定量分析，不同基因型的生物体进行代谢组学表型研究，不同生长环境及加物理、化学或生物性刺激后个体代谢产物的应答，进行代谢途径或代谢网络解析与转基因评价。

5. 动植物组织，细胞，血液，尿液等样本一般能搜到多少种代谢物？

答： 代谢物在不同的样品中，种类各异，一般植物（20 万种）、动物（2500 种）、微生物（1500 种）。在 GC-MS 平台：血液 400 左右，尿液 600 左右，动植物组织或细胞在 500 到 1000+ 不等，与生物种类有关；LC-MS 平台可以检测到的物质峰可以达到几千甚至上万，但是由于数据库的限制，目前都是做完差异分析后再做定性，一般禅意代谢物会控制在 200 个左右）。

6. 代谢组和毒理学的关系？

答： 毒理学是一门研究外源因素（化学、物理、生物因素）对生物系统的有害作用的应用学科。是一门研究化学物质对生物体的毒性反应、严重程度、发生频率和毒性作用机制的科

学，可以对毒性作用进行定性和定量评价。

代谢组学以指标分析为基础，高通量检测为手段，通过研究生物体内代谢物的种类与数量及其变化规律来阐述机体在正常生命状态及环境变化后的代谢过程。代谢组研究内容之一是不同的生长环境及加物理、化学或生物性刺激后生物体代谢产物的应答，代谢途径或代谢网络的解析。

动物的毒害与植物的毒害，植物的生物胁迫（病虫害）与非生物胁迫（旱、盐、冷、热、紫外、重金属等）都会对生物体产生刺激。因此毒理学可以作为代谢组学研究的手段，而代谢组学研究可以为毒理学研究基础的阐述提供参考，两者相辅相成，以研究生物体代谢机制。

7. 目前的主要代谢组技术有哪些？各有什么特点？

答：目前主要的代谢组技术平台主要有 NMR（核磁共振）、GC-MS（气相色谱-质谱）、LC-MS（液相色谱-质谱）。

NMR 技术具有无损伤性，无放射性，无偏向性，方法灵活，处理简单等优点，但灵敏度较低，动态范围有限；

GC-MS 技术具有高分辨率，高灵敏度，有比较标准的数据库，易于定性等优点，但需衍生化，预处理繁琐；

LC-MS 技术具有灵敏度较高，较宽动态范围，无需衍生化等优点，但标准谱图库信息不全，不易定性；

8. 应该如何选择实验平台？

答：代谢物分布广，性质差异大，单靠一种分离分析手段难以进行无偏向的全面分析，应根据样品的性质及研究目的来选择并综合利用多种技术平台。

植物代谢物主要分为初生代谢物与次生代谢物两大类。

进行初生代谢研究时，大多使用 GC-MS，

进行次生代谢研究时，大多使用 LC-MS，

LC-MS 使用的物质范围则相对更为广泛一些。



9. 代谢组可以精确描述样本内所有可能的代谢物吗？

答：目前由于分析技术上的局限性，尚未产生出一种分析技术可以精确描述样本内所有可能的化合物。

鉴于这种分析手段上的欠缺，代谢组学的最终目标还只是科学家们的理想。各大分析仪器公司也都将代谢组学目标设定为今后研发的一个重要方向，随着分析技术的逐步进步，

更精确更全谱的分析仪器将会逐步被研发出来，进而为代谢组学提供更强大的技术平台。

10. 代谢产物分析有哪几种？代谢靶标分析和全代谢组分析区别？

答： 根据研究的对象和目的不同，生物体系的代谢产物分析分为以下几个层次：

代谢物靶标分析：某一个或某几个特定组分的定性和定量分析，如某一类结构、性质相关的化合物（氨基酸、有机酸、顺二醇类）或某一代谢途径的所有中间产物进行分析，不分离鉴定具体单一组分；

代谢物指纹分析：同时对多个代谢物进行分析，不分离鉴定具体单一组分；

代谢轮廓分析：限定条件下对生物体内特定组织内的代谢产物的快速定性和半定量分析；

代谢组分析：对生物体或体内某一特定组织包含的所有代谢物的定量分析，并研究该代谢物在外界干预或病理生理条件下的动态变化规律。

11. 代谢组普筛研究一般每组多少个平行样本？为什么？

答：代谢物在不同组织样品中差异较大，为了可靠的统计学以及生物学分析，微生物与植物样本，每组最小样本数 8 个；模式动物最小样本数为 10 个；临床样本最小样本数为 30 个。

12. 代谢组生物统计学分析的方法和意义？

答： 代谢组学本质上是指从整体上研究生物体的代谢物，是对某一生物、组织或细胞中的所有低分子量代谢产物进行定性与定量分析的一门科学。

而对应的统计学分析属于统计学范畴，与代谢组学并无直接关系，但是可以通过统计学分析来判断样品的重复性、代谢物的差异以及进行遗传分析。

因此，统计学分析结合代谢物组学并无严格限制。常用的统计学分析包括独立样本 t 检验，配对样本 t 检验，F 检验，one-way ANOVA，Two-way ANOVA，PCA，PLS-DA，OPLS-DA，相关性分析，回归分析。

此外，遗传连锁分析，全基因组关联分析，候选基因的全基因组关联分析等手段已广泛应用与功能基因组研究。

13. 介绍一下代谢组用的什么仪器？不同公司之间的仪器的差异？检测灵敏度？精确度？

答： GC-MS 平台 Aglient 7890B-5977A

四级杆质谱仪，QMS；（是最常见的质谱仪器，定量能力突出，在 GC-MS 中 QMS 占绝大多数，无串极能力，定性能力不足，分辨力较低（单位分辨），存在同位素和其他 m/z 近似的离子干扰。

LC-MS 平台：AB 5600 Triple TOF

14. 质控样本 qc 的做法和意义。

答： 代谢物检测质控主要是为了保证样品检测过程的准确性，对提取或检测过程中差异较大的样品或代谢物质进行分析判断。例如，检测 100 个样品中代谢物质，在提取样品之后，从 100 个样品中各吸取部分样品，先作为一个混和样品。随后，在 100 个样品的仪器分析的过程中，可以在每 20 个样品跑样过程中插入检测一个混合样品，最后通过分析混样在前后五六次代谢检测数据稳定性，判断样品的好坏。

15. 质谱的类型与选择？

答： 色谱与质谱联用实现了从利用色谱进行物质分离到利用质谱进行物质鉴定的整个流程，且色谱，质谱的类型较为广泛。对于质谱而言，为了实现代谢物的定性与定量两个目的，应该选择不同的仪器进行综合分析。可以利用核磁共振质谱，高分辨的质谱与三重四级杆质

谱一起进行定性分析；而定量分析利用基于三重四级杆系统的质谱，如 QTrap，则更为准确。常用的高分辨质谱有 TOF-MS，Orbitrap-MS；三重四级杆系统的质谱主要为 QTrap 与 3Q 系统的质谱。

16. GC-MS 代谢组为什么要进行衍生化？

答：气相色谱仅适用于沸点低、热稳定性好的小分子物质的分离分析，对沸点高、挥发性低、热稳定性差、极性甚至极易挥发的物质往往不能直接进样分析。生物样本中含有大量的羟基、胺基、羧基、巯基等官能团的物质信息，因此采用适当的衍生化处理对于样品的分离分析往往起着十分重要的作用。样品的衍生化方法的益处有：

(1) 将一些不适合色谱分析的物质进行适当的化学处理转化成相应的挥发性衍生物，可以扩大气象色谱的测定范围；

(2) 改变同分异构化合物的色谱性能，提高和改善样品的峰形和分离度，克服载体、柱壁对高极性、低挥发样品的吸附。甚至通过特殊的衍生方法，分离手性化合物；

(3) 改善待测物质的热稳定性，以提高检测的灵敏度；

(4) 改善待测物质的质谱行为，有利于鉴定化合物的结构。

17: 在找到的代谢物列表中为什么会有名称不同保留时间不同的现象？

答：代谢物的数据库查询并不表示该离子就是这个物质，只是提供代谢物注释。根据物质分子量分解谱等信息查询，只能提供代谢物鉴定的参考，而物质的鉴定需要根据化学标准品与核磁共振质谱来分析。此外，质谱检测的时候也会有离子化与衍生化，进行物质鉴定需要综合考虑各方面因素。

18: 重名现象的两代谢物需要叠加吗？

答：不应该叠加。另外需要注意的是，重名的代谢物，需要根据保留时间、分子量、分解谱等方法来判断它们是不是同一个代谢物；如果不是，命名的时候必须予以区别，相应地在数据分析时候单独对待。

19: 如何选择检索的数据库？

答：代谢组学数据库多种多样。对于部分初生代谢物的气相质谱分析，有参考性较好的数据库；而对于液相质谱分析，目前没有固定的数据库。不同的液相质谱检测条件差异很大，对应检测的物质，在保留时间，一级分子量，二级及多级分解谱差异也很多，难以统一标准进行比较。通过多种数据库查询，比较、推断和分析将有助于解析物质结构。

20: 土壤、饲料等非常规物质代谢组的意义如何？如何去设计实验？

答：土壤中成分非常复杂，如含有营养物质、垃圾物质、动植物残留、微生物等成分。因此，对于土壤，不推荐进行代谢组学研究。不同的饲料，成分不一，也需要根据饲料的组成的复杂度来判断是否适合于代谢物分析。在适合代谢物研究的基础上，实验设计可以参考动植物代谢组研究。

21: 非模式生物进行生物信息学分析有什么难点, 该如何进行?

答: 非模式生物往往生物信息数据量少, 不太适合直接进行组学数据统计与分析, 一般建议将非模式生物的数据通过同源比对的方式映射到模式生物上, 然后再以模式生物的数据进行系统的生物信息学分析。

22: 功能未知的蛋白如何进行分析?

答: 可以通过同源比对的方式获取其关系最近的有功能的蛋白作为参考进行后续分析。

24: 同一个处理, 不同的重复之间进行非靶向代谢测定。每一个重复是一个单独的叶片、是不同株的, 物种是烟草、大田的样品, 测出来的结果差异影响大不大?

答: 不同植株同一处理, 生物学重复。样本确保一致(生长情况, 植株大小, 采样步骤等均要一致), 8个重复间的差异一般都是能接受的。

25: 如果找到的差异代谢物没有标准品, 该如何进行鉴定?

答: 一般情况下, 如果没有找到合适的标准品, 结构相近的物质也可以替代。

26: 什么是广泛靶标? 比如 TCA 三羧酸循环的一类代谢物怎么测?

答: 广泛靶向: 广泛靶向代谢组学既有非靶向代谢组学的优点也有靶向代谢组学的优点。这种广泛靶向的是不需要标准品的, 这项技术是在前期已经建立了公共数据库以外的更多的代谢物, 它的数据库更加完善, 可以检测到更多。并且这个技术使用的是三重四级杆, 这个是在使用在靶向代谢组学上的技术, 所以说它具有**靶向代谢组的优点**。①**高灵敏度**: 通过离子阱进行富集 可以检测到很多低丰度的次生代谢物, 而非靶技术无法检测到低丰度物质; ②**定性准确**: 相比非靶技术, 广靶利用分子量的同时利用二级谱进行物质定性; ③**数据库全**: 每个物种都可以自建数据库, 建立属于该物种特有的代谢库。

27: 数据库的类型有哪些?

答: GC-MS 数据库: Nist、feihn;

LC-MS 数据库: HMDB、Metlin、自建库(大连化物所-许国旺组)。

28: 非靶向代谢组学实验需要加入内标吗?

答: 一般加入 L-2-氯苯丙氨酸为内标; QC 中另加入内标共 13 种, 分别为 Carnitine C2:0-d3, Carnitine C8:0-d3, Carnitine C10:0-d3, Carnitine C16:0-d3, LPC 19:0, FFA C16:0-d3, FFA C18:0-d3, CDCA-d4, CA-d4, Trp-d5, Phe-d5, SM12:0, Choline-d4, 进行保留时间校准。

在样品中加入一定量的标准物质, 它可被色谱柱所分离, 又不受试样中其它组分峰的干扰, 只要测定内标物和待测组分的峰面积与相对响应值, 即可求出待测组分在样品中的百分含量。

内标法的关键是选择合适的内标物。内标物应是原样品中不存在的纯物质，该物质的性质应尽可能与欲测组分相近，不与被测样品起化学反应，同时要能完全溶于被测样品中。内标物的峰应尽可能接近欲测组分的峰，或位于几个欲测组分的峰中间

内标法的优点是：进样量的变化，色谱条件的微小变化对内标法定量结果的影响不大，特别是在样品前处理（如浓缩、萃取，衍生化等）前加入内标物，然后再进行前处理时，可部分补偿欲测组分在样品前处理时的损失。若要获得很高精度的结果时，可以加入数种内标物，以提高定量分析的精度。

29: 代谢组中植物样本准备中细胞破壁的方法有哪些？

答：主要有：**(1)机械法**：处理量大，速度快；例如球磨法、高压匀浆法、X-压榨法和超声波法等；**(2)酶解法**：利用水解酶破坏细胞壁中的聚合物成分；例如溶菌酶可水解细菌、放线菌细胞壁上的肽聚糖，纤维素酶可水解植物细胞壁上的纤维素，蜗牛酶可水解真菌细胞壁上的几丁质和甘露聚糖等；**(3)其他方法**：包括**细胞自溶法、渗透冲击法、反复冻融法、酸碱处理法和表面活性剂处理法等。**

30: 对于样品中存在易挥发性的物质，在保存是该如何注意才能保证实验时这些物质没有挥发掉？对于样品中存在易氧化的物质应当注意什么？

答：易挥发性的物质，收集过程中难免会有丢失，如果气体或者极易挥发可采用进样小瓶收集。含易氧化的物质的样品收集要快，尽量少暴露于空气中，避空保存，有条件可以抽真空。微生物代谢速度非常快，所以所有的步骤需要尽快完，且每个样本量一样多。

31: 不同的样品都需要多少量？

答：**细胞**: 1×10^7 (以动物细胞两种形态为例)

悬浮培养的细胞:

连同培养基转移到 15 mL 的离心管(进口)中低速离心，使细胞沉淀于离心管的底部 (1×10^7 个细胞为宜)，倒掉培养基(尽量倒干净)后可以用 PBS 快速清洗一次，再次离心，倒掉 PBS (尽量倒干净)，将离心管的尖端插入液氮中，淬灭细胞-80℃冻存，干冰邮寄，每个样品最好准备两管。

②贴壁培养的细胞:

快速倒掉培养基倒入 PBS (若用移液枪加，则靠着培养皿壁加入，以免将细胞冲起)，清洗一次，倒掉 PBS，用移液器吸尽残余的 PBS，将培养皿底部(外壁)接触液氮 10 s，淬灭细胞加入 500 微升预冷的甲醇-水 (4: 1, V/V) (需色谱纯的甲醇以及双蒸馏水)，用细胞刮刀将细胞刮下，移液枪转移至 1.5 mL 的离心管中向培养皿中加入 500 微升预冷的甲醇-水 (4: 1, V/V) 将剩余的细胞尽量全部转移至 1.5 mL 的离心管中-80℃冻存，干冰邮寄。

血清: 200ul

用离心管收集血液，37℃ (或室温) 静置 30min 或 1h 进行凝固分层。，3000rpm 4℃或常温 离心 10min，待血细胞完全沉降都管底后，取上清转至干净的离心管中；12000rpm 4℃

离心 10min, 取上清分装到 1.5mL EP 管中, 每管 0.1-0.2mL; -80℃冰箱冻存。足量干冰寄送。

血浆: 200ul

用肝素钠抗凝管(肝素、肝素钠、肝素锂)采集全血, 并尽快进行血浆分离; 采血后尽快于 3000 rpm (4℃), 离心 15 min, 待血细胞完全沉降都管底后, 吸取上层血清, 用 EP 管分装, 取上层, 每管 0.1-0.2mL 分装至 1.5mL EP 管内中; -80℃冰箱冻存。足量干冰寄送。

尿液: 1mL

晨起中段尿(临床)或晨间 1h 尿(动物)直接分装到离心管(进口)中, 每管 1mL; 添加质量体积为 1/100 (w/v) 的叠氮化钠; 收集好的尿液 10000-15000 转/分钟, 于 4℃温度下离心 10 分钟后, 吸取中层澄清的尿液, 用进口的 1.5 mL 离心管分装保存, 200 微升/管, 最好保存两管, 以保证两次实验的量。-80℃冰箱冻存, 足量干冰寄送。

唾液: 0.5 mL

禁食禁烟禁酒 1h 以上

上午 9:30-11:30 取样, 蒸馏水漱口, 将唾液外吐于无菌痰杯内(保持冰浴), 不要咳痰, 收集 5-10min, 4℃, 12000rpm, 离心 20min, 取上清, 再经 0.22um 滤膜过滤除菌, 500ul 分装, 保存于-80℃。

粪便或肠道内容物: 200mg-5g

收集到新鲜粪便或肠道内容物样本, 从粪便的不同部位挑 3 个点, 称 5g/例放到离心管中; 添加一滴(约 10uL)质量体积比为 1/100 (w/v) 的叠氮化钠(1mg/ml), 分装样本 200mg/管; 迅速放入液氮中冷冻处理至少 15min; -80℃冰箱冻存。足量干冰寄送。

常规动物组织(心、肝、脾、肺、肾、肠、脑、肌肉等): 200mg 最佳

如果组织上有血, 先用生理盐水漂洗掉; 根据具体实验设计取特定的部位, 200mg/管装入标记好的离心管中; 迅速放入液氮中冷冻处理至少 15min; -80℃冰箱冻存。足量干冰寄送。

32: 如何判断一个 GC-MS 项目结果的好坏?

答: 主要从如下方面考察: (1) 样本的代谢物的提取和检测是否得到足够多的峰; (2) 组间分群是否好。

33: TIC 图形应以平行的方式将各组放在一个图中进行峰的比较, 还是单独进行展现?

答: 代谢组学的文章中, 基本上都是单图展示, 看起来比较清楚, 当组别较多时放在一起会很混乱, 项目组别少的时候, 可以放在一起展示。

34: ChromaTOF 软件是用于色谱峰的解卷积吗? 能否用来对其他类型仪器所得数据进行分析? 进行解卷积确定代谢物峰的依据是什么? 另外, 解卷积是仅对多重峰进行处理吗? 还是将原始数据导入软件后的流程化处理? 如果有对单一色谱峰进行解卷积处理的话? 如何保证准确性?

答：所谓解卷积指对软件能将重叠的质谱图分开为“单一组分”的干净质谱图，以便更准确地定性和定量。ChromaTOF 软件的优点是可以将去噪、解卷积、定性以及峰对齐一起快速地流程化完成。对 GC-MS 的数据来说，是最方便的处理软件，其它的软件能完成其中的一两步，但是要全都完成是不行的。单一色谱峰无需进行解卷积。

35：峰与代谢物如何一一确定？以标准品还是搜库？若是搜库，通常一个色谱峰对应的化合物会有一个列表，给出可能性从大到小的排列，如何从中选择？若是选可能性最大的化合物，可能不准确；若是选其他化合物，如何保证准确性？

答：先搜库，以软件给出的默认的那个代谢物为最大可能性的代谢物；若有标品，需要进行比对才能确保准确性。

36：代谢物的鉴定结果 50%的相似度一定不可靠吗？

答：根据我们的经验，50%的相似度的定性结果未必不正确，而 70%的相似度结果未必就正确，因为 GC-MS 分析中，总会有很多共流出的峰或低响应强度的峰，这些峰的定性时软件给出的相似度就会偏低，但它的定性结果未必就不准确，这个相似度讲的只是一种可能性，为了让研究者得到更多可供选择的差异代谢物，以防我们筛选标准过严而漏掉一些研究者感兴趣的差异代谢物，我们一般会将 50%以上的结果都列上，这样研究者可以根据自己的实际情况取舍。

37：丰度是以峰面积数据，单位什么？丰度较低的怎么确定是不是背景？？

答：丰度就是代谢物的质谱响应强度，没有单位的。在用软件对原始数据进行前处理时（包括去噪音、解卷积等）设置信号与噪音的比值为 10 作为搜库的阈值，所以不会是背景噪音。

No.	Mode	Type	A	N	Models				
					R2X(cum)	R2Y(cum)	Q2(cum)	R2	Q2
A1-a2	M2	PCA-X	3	16	0.483		0.0418		
A1-a2	M3	PLS-DA	2	16	0.259	0.995	0.817		
A1-a2	M4	OPLS-DA	1+2+0	16	0.353	0.999	0.834	0.991	-0.123

PCA 模型参数：3 个主成分；R2X=0.483；Q2=0.0418；

PLS-DA 模型参数：2 个主成分；R2X=0.259；R2Y=0.995；Q2=0.817；

OPLS-DA 模型参数：3 个主成分；R2X=0.353；R2Y=0.999；Q2=0.834；

OPLS-DA 模型验证参数：R2=0.991；Q2=-0.123；

图 1

38：如上图 1 所示，相应 Type 对应的参数 (R2X, R2Y, Q2, R2, Q2) 的值在哪个范围最好？OPLS-DA 为什么会有两个 Q2？另外 cum 表示什么？

答：PCA 分析中 $R^2X > 0.4$ 为好；PLS-DA 和 OPLS-DA 分析中， R^2X 这个参数不重要了，主要是 R^2Y 和 Q^2 ，这两个值 > 0.5 为好，越接近 1 越好。OPLS-DA 中 $Q^2(\text{cum})$ ，是指建模后模型的预测能力，以大于 0.5 为宜，越接近 1 越好，cum 表示累积的意思。另外一个 Q^2 是进行模型验证，以防止随机拟合或过拟合的一个评价参数。

39：如上图 1 所示，结果中的 A, N, R^2X , R^2Y , Q^2 , R^2 , Q^2 分别代表什么意思，它们的值在什么范围比较好？

答：①A 代表建模时主成分的个数；②N 代表分析的样本数；③ R^2X 代表多元统计分析建模时，在 X 轴方向模型的解释率（或可以理解为 X 轴方向保留原始数据信息百分比的平方）；④ R^2Y 代表在 Y 轴方向模型的解释率（或可以理解为 X 轴方向保留原始数据信息百分比的平方）；⑤ Q^2 代表模型的预测率；⑥ R^2 代表模型验证时，将原始分类的 Y 矩阵、n 次不同排列的 Y 矩阵与 R^2Y 、 Q^2Y 进行线性回归，得到的回归直线与 y 轴的截距值分别为 R^2 和 Q^2 ；用来衡量模型是否过拟合。

40：PCA 得分图中为什么 A2 蓝方框和 A1 红圈会在圈外，是不是圈外的就表示结果不好，有什么方法改进？

答：这个圈圈表示 95% 的置信区间，圈外的点表示该样本的代谢谱和其他的所有分析样本的代谢谱的差异性大，一般样本数较大时会有一些离群点出现，是正常的现象，而且这两个离群点就在 95% 置信区间的周围，并未离得太远，建议保留，不需要做任何处理。

41：GC/MS 或者 LC-MS 中，圈外的点是什么原因造成的？能不能剔除掉？

答：圈外的点表示该样本的代谢谱与其他分组样本的代谢谱差别较大，出现的原因有很多（包括样本采集、储存、运输、前处理、分析等），可以剔除也可以不剔除，由于离群的并不远，为了保持数据的真实性，建议保留。

42：没有原始数据，组与组之间数据的平行性如何？如果平行性不好，如何确保检测结果的准确性？组与组之间内标峰的差异大吗？是根据内标峰面积还是峰高进行归一化处理？依据是什么？进行归一化处理用到的数据是解卷积之后的数据吗？

答：我们可以提供原始数据的，前三个应该都是想问如何保证组间差异是真实存在的而非人为因素引入的。关于组学的数据质量控制，主要看质控样本的聚类，质控样本是贯穿整个分析过程，只要质控样本聚类不错，那么数据的准确性就能保证。引起组内差异大的原因非常多，比如：个体差异、平行样本也有差异、前处理和分析过程中的引入的差异，由于有质控样本做数据质量控制，因此，前处理和分析过程中的引入的差异是可以排除的。关于归一化的：我们通常采用的是总峰面积归一化法，这也是一种非常常用的归一化方法。原始数据先进行去噪、解卷积、定性以及峰对齐后，才进行总峰面积归一化。

43: 在筛选出的 N 个差异代谢物中, 有的代谢物在 KEGG 网站里找不到 ID。在能找到 ID 的 N-M 个代谢物中有 X 种代谢物是样本中不存在的物质, 既然样本中没有这种物质, 为什么会被检测出来? (以奶牛样本为例)

答: 我们搜库是基于人和动物的代谢组学数据库 (Feinh 库) 搜的, 因此应该说这些代谢物在人和动物体内都会有的内源性小分子, 很多代谢物在 KEGG 中没有对应的 ID 号, 属于正常, KEGG 并没有把所有的小分子代谢物都囊括进去, 至于 N-M 个代谢物中有 X 种代谢物是奶牛血液中不存在的物质, 这个结论是不正确的。有许多小分子代谢物在植物和动物体内都是存在的, 植物的初级代谢物一般和动物的内源性小分子是相通的, 只有次级代谢产物不同而已。另外 Feinh 库中的代谢物就是人和动物的小分子 GC-MS 数据库, 它定性的结果应该是人和动物中都存在的小分子。

44: 代谢组学与其他组学相比, 有什么优势?

答: ①基因和蛋白质表达的微小变化会在代谢物水平得到放大; ②代谢组学的研究不需进行全基因组测序及建立大量表达序列标签的数据库; ③代谢物种类远少于基因和蛋白的数目, 每个生物体中代谢产物大约在 10^3 数量级, 而最小的细菌, 其基因组中也有几千个基因; ④生物体液的代谢物分析可反映机体系统的生理和病理状态。

45: 非靶标和靶标定量如何选择?

答: 如果没有特别关注的代谢物质建议选择非靶标, 可以检测出某一特定条件下所有的代谢产物, 并进行定性和相对定量分析, 以寻找目标差异代谢物; 如果有关注的目标代谢物建议选择靶标定量, 对特定的代谢物群进行有针对性的特异性检测。

46: 关注的目标代谢物在非靶标代谢组学检测没有检测到是什么原因?

答: 目前还没有技术及平台可以检测到全部的代谢物质, 这属于比较正常的现象; 或者已经被检测到, 但是在目前现有的数据库中不能定性, 建议使用标品在未定性的物质中进行比对。

47: 非靶标代谢组学常用的数据库有哪些?

答: NIST、Fiehn、HMDB、KEGG、METLIN、MassBank、DrugBank、Lipidmaps 等。

48: HMDB 和 METLIN 数据库有什么区别?

答: HMDB 是人类代谢组数据库 (加拿大), 该数据库不支持批量搜索, 仅限于单个代谢产物搜索, 搜索效率较低, 不支持代谢通量搜索, 代谢化合物浓度搜索等; METLIN 是 Agilent 个人物质数据库, 该数据库具有大量的 MS/MS 图谱, 而且每个化合物都有不同的碰撞能图谱, 可以清晰地找到代谢产物的碎片离子, 可以获得分子量, 化学结构式, 化学结构等信息。

49: GC-MS 和 LC-MS 检测平台各有什么优劣势?

答: 目前应用最广泛, 最有效的这两种技术可以检测包括糖、糖醇、有机酸、氨基酸、脂肪酸以及大量次生代谢物等。GC-MS 适用于分析容易气化的低极性, 低沸点的代谢物, 如: 各类挥发性化合物或者衍生化后低沸点的物质, 主要为初级代谢产物; LC-MS 不受样品挥发性和热稳定性的影响, 样品前处理非常简单, 过滤后直接进样, 可有效分析植物中丰富的次生代谢产物。

50: GC-MS 和 LC-MS 检测平台离子化方式有什么不同?

答: GC-MS 使用的是电子碰撞(EI); 而 LC-MS 使用的是电喷雾(ESI)或低频常压化学电离(APCI)。

51: 通常采用的非靶标液质联用是什么平台?

答: 本公司目前非靶标液质联用检测平台为: LC-QTOF-MS/QE focus, 色谱部分采用的是 Agilent 1290, 质谱部分采用的是 AB 6600。通常液质的一级和二级数据分开采集, AB 6600 是 DDA(数据依赖性采集方式), 对于样本中很多低丰度的小分子物质采集性不是很好。对于大的样本量(至少几百例的样本), 会把一级和二级分开采集, 采用 DIA(数据非依赖性采集方式), 可能会出现 AB 6600 和 AB 5600 联用。

52: DDA 和 DIA 数据采集方式有什么不同?

答: DDA 是数据依赖型, 只用 AB 6600 数据采集时可能会漏掉很多信息。由于采用的是 DDA 模式, 一般在采集一级质谱时会设置一个 0.5s 的时间窗口, 然后对每个 0.5s 里面出现的最高的 6 个一级峰进行提取采集, 然后去打二级质谱, 这样就导致一些峰强度比较低的物质没有被采集到, 也就不会打二级质谱。最终的定性结果里就不会出现该物质的二级定性信息; DIA 是数据非依赖型, 当采集完一级质谱后会设置一个最低强度值, 只要是一级质谱的峰强度高于该值, 就会对其进行采集并打二级质谱, 采集的数据多, 信息全, 但成本高。

53: 代谢物提取通常采用什么方法?

答: 色谱相分离通常取决于含有甲醇、水和氯仿的萃取溶液, 具有高溶解度的代谢物, 例如碳水化合物通常在甲醇、水中提取; 而具有低溶解度的代谢物, 例如膜脂和脂肪酸通常在氯仿、水中提取。如果有特别关注的化合物可以提前说明, 会在代谢物提取方法中稍作修改, 以便获得更多想要的结果。

54: 代谢组学中需要了解的质谱知识 | 扫描模式

答: **Full scan:** 全扫, 代谢组学研究中最常用的数据采集方法, 在样品采集过程中碰撞池不加能量或加很小的能量($< 5V$ 不同的仪器设置不同), 用以获得代谢物离子的一级质谱图。**Target MS/MS:** 在此工作模式下, 碰撞池施加能量, 对选定的目标离子进行二级串联, 获得二级质谱图。目标离子的选择需要手动输入, 即需要定义目标离子的质荷比以及保留时间等信息。

DDA: 全称为 data dependent acquisition 数据依赖型扫描, 此外, IDA (information dependent acquisition), auto-MS/MS 等指的也是这一扫描模式。在这种工作模式下质谱仪可以自动地在 full scan MS 和 MS/MS 采集之间进行切换, 即质谱仪可以自动的对目标离子进行碎裂, 获取二级质谱图。与 target MS/MS 不同的是, 目标离子的选择过程是自动的, 即研究者需要在样本检测之前就设定一些筛选标准, 最常见的筛选条件是设定离子强度强度阈值, 选择强度最高的几个离子 (Top-n) 进行碎裂。

DIA: 全称为 data independent acquisition 数据非依赖型扫描, 主要包括以下几种技术: all-ion fragmentation (AIF) (热电 orbitrap 系列质谱), MSall, MSe (沃特世 Q-TOF 系列质谱) 等。在此工作模式下, 碰撞池的能量在低能量和高能量之间切换, 低能量获取离子的一级质谱信息, 高能量获取离子的二级质谱图。整个过程也是自动的, 与 DDA 不同的是, 该模式不对离子做预先的筛选, 即在某一时刻所检测到的所有离子都会被高能量打碎。

SWATH: 全称为 sequential windowed acquisition of all theoretical fragment ions, 属于 DIA 技术的扩展。在此工作模式下, 扫描范围被划分为以某一固定宽度 (如 25Da) 为

间隔的一系列连续的区间，通过高速扫描来获取扫描范围内全部离子的碎片信息。

用一张图来对比一下三种获取离子碎片信息的方法（DDA，DIA，SWATH）

