

一种基于 UPLC/HRMS 的代谢组学相对定量分析方法

技术领域

本发明属于代谢组学技术领域，涉及一种基于UPLC/HRMS的、利用同位素内标的、
5 用于相对定量检测小分子代谢物的方法。

背景技术

代谢组学（Metabolomics）研究生物体系受外源性物质刺激、环境变化或遗传修饰等因素影响后产生的所有代谢应答，并检测代谢应答的全貌及其动态变化，研究对象一
10 般为相对分子质量小于1000的小分子代谢物，以高通量、大规模实验方法结合模式识别、专家系统等计算体系为特征，对生物体内所有代谢物进行定性定量分析，是系统生物学的重要组成部分。

根据研究目的不同，代谢组学可以分为四个层次：代谢物靶标分析（metabolite target analysis）、代谢轮廓分析（metabolic profiling analysis）、代谢指纹分析（metabolic fingerprint
15 analysis）和代谢组学研究（metabonomics），其中代谢物靶标分析和代谢轮廓分析用于定量检测某个或某类特征性代谢物，代谢指纹分析用于定性或半定量分析全部代谢物，代谢组学研究则用于定量分析全部代谢物，但最后一个层次目前尚未能实现。

鉴于代谢组学的研究方法主要是对代谢物进行定性定量分析，其检测手段则以化学分析领域当下的主流技术为主，包括核磁共振（NMR）、质谱（MS）、高效液相色谱（HPLC）、
20 超高效液相色谱（UPLC）等。核磁检测技术的优点是样品无需繁琐的前处理，可以实现样品无创性和代谢物种类无偏向性检测，而缺点是相比于质谱检测技术的灵敏度较低，动态范围有限，对含量较低或浓度相差较大的物质难以同时测定；另外，核磁检测技术目前偏向于定性研究，适合于鉴定未知代谢物。而质谱检测技术则具有较高的检测灵敏度，与色谱技术联用后更可以将复杂样本中的代谢物先进行色谱分离再进行质谱检测，
25 减少了基质的干扰，有利于痕量代谢物的检测。

质谱检测技术依赖于代谢物在离子源中发生电离后生成的不同质荷比的带电荷离子，在质量分析器中实现分离和聚焦，从而获得质谱图。根据联用的色谱仪器不同，色谱质谱联用技术可以分为气相色谱-质谱联用（GC-MS）、液相色谱-质谱联用（LC-MS）
30 和毛细管电泳-质谱联用（CE-MS）等。气相色谱-质谱联用技术适合于分析极性小、易挥发和热稳定性较好的代谢物，针对挥发性较低的代谢物需要衍生化预处理，而衍生化过程受水分因素影响较大，且很难反映代谢物在生物样本中真实情况。液相色谱-质谱联用

技术适合于分析中等或大极性、中低挥发性或非挥发性和热稳定性较差的代谢物。而根据质谱检测器不同，又可以分为四极杆检测器（单四极杆和三重四极杆）和高分辨率检测器（飞行时间/TOF、轨道离子阱/Orbitrap和傅里叶变换离子回旋共振/FTICR），其中四极杆质谱适合代谢物的定量检测，高分辨率质谱适合代谢物的定性分析。由于质谱检测技术对于代谢物的化学性质具有明显的偏向性，因此难以利用单独一种检测手段来同时完成定性和定量，需要多平台互补分析才能较为全面地考察代谢物的变化概况。

由于存在上述技术缺陷，目前只能围绕某单一平台开展定性或定量代谢组学检测方法。例如应用GC-MS平台来检测生物样本中的脂肪酸等，利用LC-MS/MS平台来开发靶向代谢物（包括胆汁酸、氨基酸和神经递质等）检测方法，采用高分辨质谱平台来鉴别代谢物的结构。因此，对生物样品中的代谢物进行全面、高通量、高灵敏度的同时定性定量分析仍是代谢组学研究亟需攻克的难点。

中国发明专利申请CN104297355A公开了一种基于液相色谱/质谱联用的拟靶标代谢组学分析方法，但该方法仍然需借助两种质谱平台（Q-TOF和QQQ）的基础上进行相对定量分析，浓度分析结果以当批次的质控血清（QC）经校正计算所得。Chen等公开了一种数据依赖的靶向定量代谢组学检测方法（DITQM），但该方法也是基于定性或定量两种质谱设备，首先通过高分辨率质谱获取代谢物的离子对信息，再通过三重四极杆质谱录入这些离子对，以当批次的质控血清（QC）作为校正物质来计算代谢物相对定量浓度（参见Chen Y., Zhou Z., Yang W., *et al.*, Development of a data-independent targeted metabolomics method for relative quantification using liquid chromatography coupled with tandem mass spectrometry [J], *Anal. Chem.*, **2017**, 89 (13) : 6954-6962）。尽管上述两种检测方法在一定程度上解决了定性和定量的问题，但仍需要同时使用两种质谱平台，而且均以各自实验室的质控血清（QC）样本作为计算浓度的定量校正样本，导致各实验室测得的浓度结果相互独立，无法兼容。尤其重要的是，上述采用的方法未对代谢物或差异代谢物进行结构鉴定（或注释），并不清楚是何种代谢物发生浓度变化，因而对代谢组学研究的意义不大，未从根本上解决代谢组学的同时定性和定量技术问题。

发明内容

发明要解决的问题

针对上述技术问题，本发明的目的在于建立一种能够基于一个平台对代谢组学样本实现同时定性和相对定量分析的方法。

用于解决问题的方案

本发明提供了一种基于UPLC/HRMS的代谢组学相对定量分析方法，其包括下列步骤：

- 1) 基于多种同位素内标，配制同位素内标混合溶液；
- 2) 基于代谢组学样本，确定与之匹配的相对定量校正样本，并利用所述相对定量校正样本和步骤1)中所述同位素内标混合溶液，配制一系列浓度梯度的相对定量标准曲线校正溶液；
- 3) 利用步骤2)中所述代谢组学样本和步骤1)中所述同位素内标混合溶液，配制代谢组学样本溶液；
- 4) 利用UPLC/HRMS平台，采集步骤3)中所述代谢组学样本溶液和步骤2)中所述一系列浓度梯度的相对定量标准曲线校正溶液的原始质谱数据；
- 5) 基于步骤4)中所述原始质谱数据，获得一级质谱转置数据和二级质谱转置数据，并基于所述一级质谱转置数据，获得包括多种一级变量信息的解卷积结果；
- 6) 结合步骤5)中所述一级质谱转置数据的解卷积结果与所述二级质谱转置数据，并参比单个同位素内标的一级变量信息和二级变量信息，鉴定同位素内标，并选择用于线性拟合的最优同位素内标；
- 7) 利用步骤2)中所述一系列浓度梯度的相对定量标准曲线校正溶液的浓度及其在步骤5)中获得的一级变量信息进行线性拟合，获得线性方程；
- 8) 基于步骤7)中所述线性方程，得到步骤3)中所述代谢组学样本溶液的相对定量结果，并完成主成分分析和一级变量信息的差异代谢物分析；和
- 9) 结合步骤5)中所述一级质谱转置数据的解卷积结果与所述二级质谱转置数据，完成代谢物鉴定和差异代谢物鉴定。

优选地，在上述方法中，步骤1)中的同位素内标混合溶液通过下列方法制得：先取适量的多种同位素内标，分别加入溶剂，配制成一定浓度的单一同位素内标母液，再分别取适量的各种单一同位素内标母液，混合并加入溶剂，得到同位素内标混合溶液。

优选地，在上述方法中，步骤2)中的代谢组学样本为血清或血浆样本，相对定量校正样本为NIST血清。

优选地，在上述方法中，步骤2)中的一系列浓度梯度的相对定量标准曲线校正溶液通过下列方法制得：取一系列体积的相对定量校正样本，分别加入同位素内标和蛋白沉淀试剂，混合，离心，取上清液，浓缩，向残渣中加入复溶溶剂，混合，取上清液，得到一系列浓度梯度的相对定量标准曲线校正溶液。

优选地，在上述方法中，步骤3)中的代谢组学样本溶液通过下列方法制得：取代

谢组学样本，分别加入同位素内标和蛋白沉淀试剂，混合，离心，取上清液，浓缩，向残渣中加入复溶溶剂，混合，取上清液，得到代谢组学样本溶液。

优选地，在上述方法中，步骤5)中的一级质谱转置数据和二级质谱转置数据通过数据转置的方式获得，该数据转置通过软件或算法平台来完成。

- 5 优选地，在上述方法中，步骤5)中的解卷积结果通过解卷积的方式获得，该解卷积通过软件或算法平台来完成。

优选地，在上述方法中，步骤6)中的同位素内标鉴定通过自研算法来完成。

优选地，在上述方法中，步骤6)中的最优同位素内标通过自研算法来选择。

优选地，在上述方法中，步骤8)中的主成分分析通过自适应换算来完成。

- 10 优选地，在上述方法中，步骤8)中的差异代谢物分析通过多元统计分析来完成。

优选地，在上述方法中，步骤9)中的代谢物鉴定和差异代谢物鉴定通过软件或算法平台来完成。

发明的效果

与现有技术相比，本发明的代谢组学相对定量分析方法具有如下有益效果：

- 15 1) 本发明的研究思路不同于以往的研究模式，无需先从高分辨质谱数据中提取多反应监测模式(MRM)离子对，再转移到三重四极杆质谱对这些离子对进行定量扫描，而是直接从高分辨率质谱数据中先提取质谱一级变量信息，以同位素内标和NIST血清样本同时作为校正物质来计算代谢组学样本中代谢物的相对浓度值，通过一级分析获得一级变量的差异代谢物，通过二级分析得到的代谢物全鉴定表格，特别是明确物质成分或
- 20 结构的差异代谢物；

2) 本发明的分析方法无需同时使用两种质谱平台(四极杆质谱和高分辨质谱)，仅用一种(高分辨质谱)即可满足定性和定量的双重需求，成本较低，操作简便，适用性广；

- 25 3) 本发明的分析方法同时以多种同位素内标和NIST血清样本作为校正物质，使得定量结果更为准确，并且使得各个实验室之间的结果具有兼容性(校准物质均为NIST血清样本)，解决了现有的代谢组学数据存在孤岛效应的问题；

4) 代谢组学样本的预处理简单易行，无目标代谢物类型的歧视效应，适合各种类型代谢物的定性和定量研究；

- 30 5) 相比于传统代谢组学研究，本发明的分析方法依赖于计算机算法脚本，可以快速鉴定代谢物结构，准确率较高，避免了人为筛选的主观性。

附图说明

图1示出了基于UPLC/HRMS的代谢组学相对定量分析方法建立流程图。

图2示出了人血清的正离子模式基峰离子流（BPI）色谱图。

图3示出了人血清的负离子模式基峰离子流（BPI）色谱图。

5 图4示出了人血清样本在正离子模式下的主成分分析（PCA）得分图。

图5示出了人血清样本在负离子模式下的主成分分析（PCA）得分图。

图6示出了采用代谢组学相对定量分析方法和靶向绝对定量分析方法的人血清样本中胆碱、甜菜碱、左旋肉碱和肌酐酸4种代谢物浓度的相关散点图。

10 具体实施方式

[术语定义]

除非另有说明，本文中所使用的术语“内标”（或“内部标准”）是指一种在定量分析过程中添加于供试样品内的适当的纯品，用于计算待测组分的含量。

15 除非另有说明，本文中所使用的术语“同位素”是指隶属同一元素的、具有相同质子数和不同中子数的不同核素之间的互称。例如，氢元素包含氕（H或 ^1H ）、氘（D或 ^2H ）和氚（T或 ^3H ）三种同位素，碳元素包含碳12（ ^{12}C ）、碳13（ ^{13}C ）和碳14（ ^{14}C ）三种同位素，氮元素包含氮14（ ^{14}N ）和氮15（ ^{15}N ）两种同位素，氧元素包含氧16（ ^{16}O ）、氧17（ ^{17}O ）和氧18（ ^{18}O ）三种同位素。

20 除非另有说明，本文中所使用的术语“同位素内标”是指结构中包含的至少一种丰度较大的同位素被替换为丰度较小的同位素之后形成的稳定的内标物质。例如，在作为内标的氯化胆碱的分子结构中，若三个甲基所包含的九个氢原子（或氕原子）全部被氘原子替换，则可得到相应的同位素内标—氯化胆碱-三甲基-d9。又如，在作为内标的L-缬氨酸的分子结构中，若羧基所包含的碳12原子被碳13原子替换，则可得到相应的同位素内标—L-缬氨酸-1- ^{13}C 。按照代谢功能不同，本发明中的同位素内标可分成碳水化合物类（carbohydrate）（如D-果糖-1- ^{13}C ）、能量代谢类（energy）（如5'-三磷酸腺苷- $^{15}\text{N}_5$ ）、核苷酸类（nucleotide）（如尿苷-2- ^{13}C ）、氨基酸类（amino acid）（如L-缬氨酸-1- ^{13}C ）、辅因子和维生素类（cofactor and vitamin）（如维生素B3-d4）等多种类别。

30 除非另有说明，本文中所使用的术语“NIST血清”是指由美国国家标准与技术研究院（National Institute of Standards and Technology，简称NIST）提供或售出的冷冻人血清（Frozen Human Serum，简称FHS）。

除非另有说明，本文中所使用的术语“蛋白沉淀试剂”是指一种添加于生物样品内，

用于使样品中的蛋白质因变性而沉淀析出的试剂。常见的蛋白沉淀试剂包括但不限于盐类（如硫酸铵、硫酸钠、氯化钠等）、有机溶剂类（如甲醇、乙腈、丙酮等）等。

除非另有说明，本文中所使用的术语“代谢组学样本”是指用于代谢组学研究的生物样本。常见的代谢组学样本包括但不限于动物样本（如人或动物的血液、尿液、粪便、
5 唾液、毛发、细胞等）、植物样本（如植物的根、茎、叶、花、果实、种子等）、微生物样本（如微生物的细胞、孢子、发酵液、培养液等）、亚细胞结构样本（如细胞器的线粒体、外泌体、囊泡等）等。

除非另有说明，本文中所使用的术语“质量控制样本”（或“QC样本”）是指将实验样本取一定量混合后得到的混合样本，目的是为了监控实验过程中仪器采集数据的稳定性和
10 和大致查看实验样本的色谱峰概况以及非血清或血浆样本相对定量检测的校准。

除非另有说明，本文中所使用的术语“一级质谱全扫描加数据依赖的二级质谱扫描模式”（或“full MS-ddMS2”）是指仪器在采集一级质谱数据的同时碎裂该离子，以生成二级质谱碎片的模式，可以同时获得一级和二级质谱数据，优选在规定时间内信号强度前3
15 离子进行二级碎裂，循环至下一时间周期首先排除已选信号强度前3离子，筛选另外信号强度前3离子进行再次二级碎裂，直至时间周期结束。

除非另有说明，本文中所使用的术语“转置数据”是指将不同的质谱仪器产生的原始质谱数据转化为标准化、可进行文本处理的质谱数据；术语“一级质谱转置数据”是指包含检测目标物的一级分子信息的质谱转置数据；术语“二级质谱转置数据”是指包含检测
20 目标物的二级分子信息的质谱转置数据；相应地，术语“数值转置”是指将原始质谱数据转化成标准化、可进行文本处理的数据的过程。

除非另有说明，本文中所使用的术语“解卷积”是指从一级质谱转置数据中分离并构建出每一个代谢物母离子的信息，包括质荷比、保留时间和对应的色谱峰信息等。

除非另有说明，本文中所使用的术语“一级变量”是指仪器采集到所有一级质谱数据（母离子）的质荷比和对应的色谱峰信息。在代谢组学领域中，常见的一级变量包括但
25 不限于一级变量编号、母离子质荷比、保留时间、峰面积值等。

除非另有说明，本文中所使用的术语“结构鉴定”（或“代谢物鉴定”）是指将色谱峰的二级质谱碎片（母离子碎片）比对公共数据库（例如KEGG）中的二级质谱碎片并且具有对应的一级质谱数据（母离子）获得的物质结构推导过程。结构鉴定后的物质会匹
30 配相应公共数据库的编号。

除非另有说明，本文中所使用的术语“相对定量浓度”（或“相对定量结果”）是指通过NIST血清样本中代谢物的浓度和代谢物峰面积与同位素内标峰面积的比值线性拟合后

获得线性方程，然后将代谢组学样本中代谢物与同位素内标的峰面积比值（或一级变量）代入线性方程，计算得到的以NIST血清样本物质浓度为基数的相对浓度。

除非另有说明，本文中所使用的术语“一级分析”是指根据一级变量的数据进行统计分析的过程；术语“主成分分析”是指代谢物变量按一定的权重通过线性组合后产生新的特征变量，通过主要新变量（主成分）对各组数据进行归类的过程；术语“自适应换算”是指对一级变量的数据进行权重换算的数据均一化过程；术语“差异代谢物分析”（或“代谢物一级变量的差异分析”）是指利用统计学的差异分析方法对代谢物一级变量的信息进行组间差异分析；术语“多元统计分析”是指在多个对象和多个指标互相关联的情况下分析它们的统计规律。

除非另有说明，本文中所使用的术语“二级分析”是指在一级分析的基础上，结合二级质谱数据进行统计分析的过程；术语“差异代谢物鉴定”是指根据一级分析的差异代谢物进行结构鉴定。

具体而言，本发明公开了一种基于UPLC/HRMS的代谢组学相对定量分析方法。该方法包括下列步骤：

- 1) 基于多种同位素内标，配制同位素内标混合溶液；
- 2) 基于代谢组学样本，确定与之匹配的相对定量校正样本，并利用所述相对定量校正样本和步骤1)中所述同位素内标混合溶液，配制一系列浓度梯度的相对定量标准曲线校正溶液；
- 3) 利用步骤2)中所述代谢组学样本和步骤1)中所述同位素内标混合溶液，配制代谢组学样本溶液；
- 4) 利用UPLC/HRMS平台，采集步骤3)中所述代谢组学样本溶液和步骤2)中所述一系列浓度梯度的相对定量标准曲线校正溶液的原始质谱数据；
- 5) 基于步骤4)中所述原始质谱数据，获得一级质谱转置数据和二级质谱转置数据，并基于所述一级质谱转置数据，获得包括多种一级变量信息的解卷积结果；
- 6) 结合步骤5)中所述一级质谱转置数据的解卷积结果与所述二级质谱转置数据，并参比单个同位素内标的一级变量和二级变量信息，鉴定同位素内标，并选择用于线性拟合的最优同位素内标；
- 7) 利用步骤2)中所述一系列浓度梯度的相对定量标准曲线校正溶液的浓度及其在步骤5)中获得的一级变量信息进行线性拟合，获得线性方程；
- 8) 基于步骤7)中所述线性方程，得到步骤3)中所述代谢组学样本溶液的相对定量结果，并完成主成分分析和一级变量信息的差异代谢物分析；和

9) 结合步骤5)中所述一级质谱转置数据的解卷积结果与所述二级质谱转置数据,完成代谢物鉴定和差异代谢物鉴定。

在一些优选的实施方案中,步骤1)中的同位素内标混合溶液通过下列方法制得:先取适量的若干种(例如9种)同位素内标,分别加入溶剂(例如水或甲醇),配制成一定浓度(例如1~15 mg/mL)的单一同位素内标母液,再分别取适量的各种单一同位素内标母液,混合并加入溶剂(例如水或甲醇),得到同位素内标混合溶液(同位素内标的数量、种类和单一同位素内标母液的浓度可根据代谢组学样本中代谢物的具体类型和实际浓度响应酌情调整)。

在一些优选的实施方案中,步骤2)中的代谢组学样本为血清或血浆样本,此时以NIST血清作为相对定量校正样本。而在其他的实施方案中,步骤2)中的代谢组学样本为其他样本(例如动物或植物的细胞、组织、器官或代谢物等),此时以质量控制样本(通过将适量的多个代谢组学样本混合而制得)作为相对定量校正样本。

在一些优选的实施方案中,步骤2)中的一系列浓度梯度的相对定量标准曲线校正溶液通过下列方法制得:取一系列(例如7个)体积的相对定量校正样本(基于代谢组学样本的具体种类确定,其体积可根据实际需要酌情调整),分别加入同位素内标和蛋白沉淀试剂(例如甲醇),混合,离心,取上清液,浓缩,向残渣中加入复溶溶剂(例如80%甲醇水(v/v)),混合,取上清液,得到一系列浓度梯度的相对定量标准曲线校正溶液。

在一些优选的实施方案中,步骤3)中的代谢组学样本溶液通过下列方法制得:取代谢组学样本(样本的数量、种类以及单个样本的取样量均可根据实际需要酌情调整;另外,根据样本的性质或种类不同,可使用移液器量取或天平称取),分别加入同位素内标和蛋白沉淀试剂(例如甲醇),混合,离心,取上清液,浓缩,向残渣中加入复溶溶剂(例如80%甲醇水(v/v)),混合,取上清液,得到代谢组学样本溶液。

在一些优选的实施方案中,步骤4)中的原始质谱数据需要借助超高效液相色谱仪串联高分辨质谱仪(简称“UPLC/HRMS平台”)并且在一定的色谱和质谱条件下来完成采集。例如,在色谱条件中,色谱柱可以选用十八烷基硅烷键合硅胶柱(例如Waters ACQUITY UPLC BEH C18色谱柱),流动相可以选用二元流动相体系(例如在正离子模式下的甲酸水溶液/甲酸乙腈溶液,又如在负离子模式下的甲酸铵水溶液/乙腈),洗脱方式可以选用梯度洗脱,流速、进样量和柱温等条件可根据待测物质的实际色谱行为酌情调整;在质谱条件中,数据采集方式可以选用full MS-ddMS2,离子源可以选用软电离离子源(例如电喷雾离子源/ESI),喷雾电压、鞘气/辅助气压力、毛细管温度、碰撞电压

和扫描分辨率及范围等条件可根据待测物质的实际色谱行为酌情调整。

在一些优选的实施方案中，步骤5)中的一级质谱转置数据和二级质谱转置数据通过数据转置的方式获得。该数据转置过程通常需要借助软件或算法平台来完成，例如开源的ProteoWizard软件。

- 5 在一些优选的实施方案中，步骤5)中的解卷积结果通过解卷积的方式获得。该解卷积过程通常需要借助软件或算法平台来完成，例如BioDeep代谢组学云分析平台（简称BioDeep云平台）或开源R软件（<http://www.bioconductor.org/>）。

- 10 在一些优选的实施方案中，步骤6)中的同位素内标鉴定通过自研算法来完成。该自研算法通过比较一级质谱转置数据的解卷积结果和二级质谱转置数据（例如，首先获取单个同位素内标的rt（保留时间）、mz（母离子质荷比）和ms2（二级碎片），然后参比代谢组学样本溶液、质量控制样本溶液和相对定量标准曲线校正溶液中单个同位素内标的rt、mz和ms2，确保对应mz的误差范围在10 ppm以内，对应rt的差值最小，ms2能够完全匹配），完成同位素内标鉴定。

- 15 在一些优选的实施方案中，步骤6)中的最优同位素内标通过下列方法选择：将NIST血清样本（或相对定量校正样本）中一级变量峰面积与同位素内标混合溶液中每种内标的一级变量峰面积进行比值，再与NIST血清样本（或相对定量校正样本）浓度线性拟合后获得相关系数r，优选r值最大以及NIST血清样本（或相对定量校正样本）中一级变量的浓度范围介于0.05P~5P的同位素内标。

- 20 在一些优选的实施方案中，步骤8)中的主成分分析通过自适应换算来完成。该自适应换算是把原始的观察数值减去每组数据的均值，并除以标准差后得到的数值。

在一些优选的实施方案中，步骤8)中的差异代谢物分析通过多元统计分析来完成。该多元统计分析通过主成分分析、t-test检验和VIP（空间投影重要性或者差异权重贡献值）计算等多元统计分析方法，对差异代谢物进行筛选，优选t-test检验结果P value $\leq$ 0.05，并同时满足VIP $\geq$ 1的差异代谢物。

- 25 在一些优选的实施方案中，步骤9)中的代谢物鉴定和差异代谢物鉴定借助软件或算法平台来完成，例如BioDeep代谢组学云分析平台（简称BioDeep云平台）。

以下将通过具体实施例来进一步说明本发明的相对定量代谢组学分析方法。除非另有说明，实施例中所使用的材料、试剂、仪器、软件等均可通过常规商业手段获得。

- 30 **实施例：**基于UPLC/HRMS的健康人和脑梗病人血清代谢组学相对定量分析方法的建立。

建立基于UPLC/HRMS的健康人和脑梗病人血清代谢组学半定量分析方法的流程如

图1所示, 具体实施步骤如下:

1、内标的制备:

1.1 制备同位素内标混合溶液:

取适量的D-果糖-1-13C (F-13C)、L-缬氨酸-1-13C (Val-13C)、氯化胆碱-三甲基
5 -d9 (Choline-d9)、5'-三磷酸腺苷-15N5 (ATP-15N5)、L-苯丙氨酸-1-13C (Phe-13C)、
尿苷-2-13C(U-2-13C)、维生素B3-d4(VB3-d4)、左旋肉碱盐酸盐-甲基-d3(L-carnitine-d3)
和甜菜碱盐酸盐-三甲基-d9(betaine-d9)等9种同位素内标, 分别加入水或甲醇配制成1~15
mg/mL的单一同位素内标母液, 再分别取适量的9种单一同位素内标母液, 混合并加入适
量的水或甲醇, 配制成9种同位素内标的浓度范围均为1~50 µg/mL的同位素内标混合溶液。

10 1.2 制备NIST血清标准曲线校正溶液:

用移液器量取NIST血清 (Lot. No.: SRM 909c) (即图1中的NIST血清样本), 共计
5、10、50、100、200、300和500 µL等7个梯度, 分别加入100 µL的第1.1项中制备的同位
素内标混合溶液和3倍体积的预先于-20℃冷冻的甲醇进行蛋白沉淀, 充分混合, 于12000
rpm和4℃环境下离心10 min, 取全部上清, 于真空浓缩干燥器中浓缩, 向残渣中加入150
15 µL的预先于-20℃冷冻的80%甲醇/水混合物 (v/v) 作为复溶溶剂, 充分混合后, 取上清
液作为NIST血清标准曲线校正溶液。

2、样本的制备:

制备健康人和脑梗病人血清代谢组学样本溶液和质量控制样本溶液:

选择10个健康人血清样本和10个脑梗病人血清样本作为代谢组学样本 (共计20个);
20 从每个代谢组学样本中各取20 µL (共计400 µL), 充分混匀后作为质量控制样本 (即图
1中的QC样本, 用于方法最终的归一化处理)。用移液器分别量取20个代谢组学样本和4
个QC样本各100 µL, 分别加入400 µL的预先于-20℃冷冻的甲醇 (蛋白沉淀试剂) 和100 µL
的第1.1项中制备的同位素内标混合溶液, 充分混匀后, 于12000 rpm和4℃环境下离心10
min; 取全部上清, 于真空浓缩干燥器中浓缩, 向残渣中加入150 µL的预先于-20℃冷冻
25 的80%甲醇/水混合物 (v/v) 作为复溶溶剂, 充分混合后, 取上清液作为健康人和脑梗病
人代谢组学样本溶液和质量控制样本溶液。

3、仪器及其条件参数的选定:

3.1 选定仪器:

Thermo Scientific Ultimate 3000液相色谱仪串联Thermo Scientific Q-Exactive质谱仪。

30 3.2 选定液相色谱条件:

色谱柱: Waters ACQUITY UPLC BEH C18色谱柱 (100 mm × 2.1 mm, 1.7 µm);

流动相：在正离子模式下，以0.1%甲酸水溶液（每1 L水溶液中含1 mL甲酸）作为流动相A，以0.1%甲酸乙腈溶液（每1 L乙腈溶液中含1 mL甲酸）作为流动相B；在负离子模式下，以5 mM甲酸铵水溶液（每1 L水溶液中含5 mM甲酸铵）作为流动相A，以乙腈作为流动相B；

- 5 洗脱方式：采用梯度洗脱，具体程序如下：0~1 min, 2%流动相A；1~9 min, 2%~50%流动相A；9~12 min, 50%~98%流动相A；12~13.5 min, 98%流动相A；13.5~14 min, 98%~2%流动相A；14~20 min, 2%流动相A；

流速：0.25 mL/min；

进样量：2 μ L；

- 10 柱温：40℃。

3.3 选定质谱条件：

数据采集方式：采用一级质谱全扫描加数据依赖的二级质谱扫描（full MS-ddMS2）模式测定，同时获得一级和二级质谱数据；

离子源：电喷雾离子源（ESI）；

- 15 喷雾电压：在正离子模式下为3.5 kV，在负离子模式下为2.5 kV；

鞘气压力：30 arb；

辅助气压力：10 arb；

毛细管温度：325℃；

全扫描分辨率为70000 FWHM，质谱扫描范围为m/z 81~1000；

- 20 采用高能诱导裂解技术（HCD）进行二级裂解，碰撞电压为30 eV，同时动态排除无必要的二级质谱信息。

4、数据处理：

4.1 原始质谱数据转置：

- 25 分别在正、负离子模式下获取包括第2项中制备的20个健康人和脑梗病人血清代谢组学样本溶液和4个QC样本溶液以及第1.2项中制备的7个NIST血清标准曲线校正溶液（共计31个）的原始质谱数据的*.raw文件。原始数据文件经ProteoWizard软件（开源软件，可以通过<http://proteowizard.sourceforge.net/>获取）进行数据转置，形成分别包括在正、负离子模式下的一级质谱转置数据和二级质谱转置数据的*.mzXML文件。

4.2 转置后数据解卷积：

- 30 第4.1项中形成的包括在正、负离子模式下的一级质谱转置数据的*.mzXML文件借助biodeep云平台（也可以通过<http://www.bioconductor.org/>下载相应的开源软件和脚本）

进行解卷积，在正离子模式下得到13128个一级变量，在负离子模式下得到17272个一级变量，对应输出作为一级解卷积结果的peaktable_pos.xlsx和peaktable_neg.xlsx文件，两个文件中均包括xc_ms id（一级变量编号）、mz（母离子质荷比）、rt（保留时间）、intensity（峰面积）等一级变量信息。

5 4.3 同位素内标鉴定和优选：

第4.2项中输出的peaktable_pos.xlsx或peaktable_neg.xlsx文件分别与第4.1项中形成的二级质谱转置数据结合，借助自研算法并参比单个同位素内标的rt（保留时间）、mz（母离子质荷比）和ms2（二级碎片），以确保人血清代谢组学样本、QC样本和NIST血清样本中mz的误差范围在10 ppm以内，保留时间与参比同位素内标保留时间的差值最小，二级碎片能完全匹配上。在正离子模式下鉴定到内标物质为Val-13C、Choline-d9、Phe-13C、L-carnitine-d3和betaine-d9等5种同位素内标，在负离子模式下鉴定到F-13C、Val-13C、Phe-13C、U-2-13C和VB3-d4等5种同位素内标，对应输出IS.xlsx文件，包括IS xc_ms id（同位素内标一级变量编号）、IS_name（同位素内标名称）、IS mz（同位素内标母离子质荷比）、rt（同位素内标保留时间）、intensity（NIST血清样本中含有的同位素内标峰面积）等一级变量信息。

利用自研算法优选合适的同位素内标，具体程序如下：第4.2项中输出的peaktable_pos.xlsx或peaktable_neg.xlsx文件中NIST血清样本一级变量intensity（峰面积）与正离子模式下鉴定到的Val-13C、Choline-d9、Phe-13C、L-carnitine-d3和betaine-d9等5种同位素内标（负离子模式为F-13C、Val-13C、Phe-13C、U-2-13C和VB3-d4等5种同位素内标）峰面积的比值与7个NIST血清标准曲线校正浓度分别进行线性拟合，每个一级变量都会获得5个线性方程和相关系数r，优选相关系数r值最大，其次为一级变量的浓度范围介于0.05P~5P的同位素内标。

4.4 NIST血清样本中一级变量线性拟合：

将第4.2项中得到的NIST血清标准曲线校正溶液中代谢物的峰面积按照从低到高的顺序进行排序，以100 μ L的NIST血清作为浓度P，5、10、50、100、200、300和500 μ L的NIST血清的相应浓度分别为0.05P、0.1P、0.5P、P、2P、3P和5P，该浓度顺序作为线性拟合的浓度点。再以NIST血清标准曲线校正溶液中代谢物的峰面积和每种优选同位素内标的峰面积的比值与上述7个浓度点进行线性拟合，计算得到线性方程和相关系数r，优选相关系数r最大，其次为一级变量的浓度范围介于0.05P~5P的同位素内标物质。在正离子模式下，相关系数r大于0.80的一级变量有8558个；在负离子模式下，相关系数r大于0.80的一级变量有5151个。

4.5 计算代谢组学样本中一级变量的定量浓度：

将4.2项中得到的代谢组学样本溶液中代谢物的峰面积与NIST血清标准曲线校正溶液中优选的优选同位素内标的峰面积的比值代入第4.4项中得到的线性方程，计算得到包括代谢组学样本中代谢物的相对定量浓度的 biodeepProfiler_pos.xlsx 或 biodeepProfiler_neg.xlsx文件（以P为浓度单位）。

4.6 一级质谱数据分析：

将第4.5项中得到的biodeepProfiler_pos.xlsx或biodeepProfiler_neg.xlsx文件分别选择数据自适应（UV）换算处理，获取在正、负离子模式下的基峰离子流（BPI）色谱图（如图2和图3所示）和主成分分析得分图（如图4和图5所示）；在多元统计分析中，选择P value $\leq$ 0.05, VIP $\geq$ 1的筛选阈值，对特定的代谢组学样本进行代谢物一级变量的差异分析。优选地，还可以利用质量控制样本进行归一化处理。

4.7 二级质谱数据分析-代谢物鉴定：

将第4.2项中得到的解卷积后的一级变量和第4.1项中形成的二级质谱转置数据结合，进行代谢物鉴定，获得包括代谢物鉴定结果的biodeepMSMS_union.xlsx文件，合并正离子模式和负离子模式代谢物，共筛选到304个代谢物。

4.8 二级质谱数据分析-差异代谢物鉴定：

将第4.7项中筛选到的代谢物与第4.6项中获取的一级变量差异代谢物分析结果进行整合，共得到33个能够明确物质结构的差异代谢物。

5、代谢组学相对定量分析与靶向绝对定量分析的比较：

选取胆碱、甜菜碱、左旋肉碱和肌酐酸等4种胆碱类似物作为代谢组学相对定量分析与靶向绝对定量分析结果比较的代谢物。

5.1 靶向绝对定量分析：

5.1.1 制备靶向绝对定量分析样本：

靶向绝对定量分析样本与代谢组学相对定量分析样本一致，在制备代谢组学样本中分别再平行制备一份，样本包括10个健康人和10个脑梗病人血清，7个NIST血清样本和4个QC样本，样本制备过程同第1项和第2项。

5.1.2 靶向绝对定量分析方法：

5.1.2.1 选定仪器：

Waters ACQUITY UPLC液相色谱仪串联AB 4000质谱仪。

5.1.2.2 选定液相色谱条件：

色谱柱：ACQUITY UPLC® BEH HILIC色谱柱（100 mm $\times$ 2.1 mm, 1.7 μ m）；

流动相：以乙腈作为流动相A，以0.1%甲酸和10 mM甲酸铵水溶液（每1 L水溶液中分别含1 mL甲酸和10 mM甲酸铵）作为流动相B；

洗脱方式：采用梯度洗脱，具体程序如下：0~1 min，80%流动相A；1~2 min，80~70%流动相A；2~2.5 min，70%流动相A；2.5~3 min，70~50%流动相A；3~3.5 min，50%流动相A；3.5~4 min，50~80%流动相A；4~6 min，80%流动相A；

流速：0.40 mL/min；

进样量：5 μ L；

柱温：40℃。

5.1.2.3 选定质谱条件：

10 数据采集方式：采用多重反应监测（MRM）进行扫描和数据采集；

离子源：电喷雾离子源（ESI）；

喷雾电压：在正离子模式下为5.0 kV；

离子源温度：500℃；

碰撞气：6 psi；

15 气帘气：30 psi；

雾化气和辅助气：50 psi；

用于定量分析的离子对分别为胆碱（Choline）（104.075>44.9）、甜菜碱（Betaine）（118.080>41.9）、左旋肉碱（L-Carnitine）（162.023>59.9）和肌酸酐（Creatinine）（114.025>86.1），同位素内标采集离子对分别为Choline-d9（113.030>69.0）、betaine-d9（127.064>68.0）和L-carnitine-d3（165.105>103.1），其中肌酸酐所对应的内标同为L-carnitine-d3。

5.1.3 靶向绝对定量数据处理：

采用自研算法，将10个健康人和10个脑梗病人血清样本、7个NIST血清样本和4个QC样本中胆碱、甜菜碱、左旋肉碱、肌酸酐、Choline-d9、betaine-d9和L-carnitine-d3的峰面积数值导出。以NIST血清样本浓度0.05P、0.1P、0.5P、P、2P、3P和5P作为自变量梯度浓度点，以NIST血清样本中胆碱、甜菜碱、左旋肉碱和肌酸酐的峰面积和对应同位素内标的峰面积的比值作为因变量，自变量和因变量进行线性拟合，得到线性方程和相关系数r。将代谢组学样本中胆碱、甜菜碱、左旋肉碱和肌酸酐的峰面积和对应同位素内标的峰面积的比值代入线性方程，计算得到代谢组学样本中代谢物的绝对定量浓度。

30 5.2 代谢组学相对定量分析与靶向绝对定量分析的结果比较：

5.2.1 代谢组学相对定量分析是以所有的代谢物为研究对象，为获取验证胆碱、甜

菜碱、左旋肉碱、肌酸酐这4种代谢物的相对定量分析结果，将第4.5项和第4.7项中获得的结果整合分析，得到这4种代谢物的相对定量浓度。

5.2.2 将第5.1.3项中得到的10个健康人和10个脑梗病人血清样本中4种代谢物的绝对定量浓度与第5.2.1项中得到的相对定量浓度进行相关分析，得到如图6所示的散点图，其中A~D依次为通过两种检测方法得到的胆碱、甜菜碱、左旋肉碱和肌酸酐浓度的相关散点图（浓度单位为P），两种方法检测到的浓度结果相关系数r均大于0.99，检测结果相关程度高，方法无区别。

5.2.3 进一步将第5.1.3项中得到的10个健康人和10个脑梗病人血清样本中4种代谢物的绝对定量浓度与第5.2.1项中得到的相对定量浓度的相对标准偏差（RSD，%）和双侧配对检验结果示于表1。如表1中所示，健康人和脑梗病人组两种检测结果浓度的RSD值无显著性差异（P value大于0.05），说明两种检测结果类似，本发明的代谢组学相对定量分析方法的准确性较高。同时，本发明的代谢组学相对定量分析方法还能基于同一个分析测试平台同时实现代谢物的定性分析。

表1. 通过两种检测方法所得的4种代谢物定量浓度的RSD值比较结果

代谢物	RSD（%）			
	健康人（10 人）		脑梗病人（10 人）	
	绝对定量	相对定量	绝对定量	相对定量
Choline	18.11	16.99	24.02	22.07
Betaine	21.93	23.55	15.34	19.90
L-Carnitine	20.48	20.78	11.96	12.09
Creatinine	10.78	11.04	31.39	30.86
P value	0.6700		0.7197	