

# Recent Development of GC-MS-Based Analytical Procedures of Metabolomics in Screening of Lung Cancer Biomarkers

Caixia Li, Peipei Zhang, Lichao Wang, Huijun Liu, Zaifa Pan, Lili Wang\*

College of Chemical Engineering, Zhejiang University of Technology, Hangzhou  
Email: [lili\\_wang@zjut.edu.cn](mailto:lili_wang@zjut.edu.cn)

Received: Mar. 10<sup>th</sup>, 2014; revised: Mar. 18<sup>th</sup>, 2014; accepted: Mar. 25<sup>th</sup>, 2014

Copyright © 2014 by authors and Hans Publishers Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The complete analytical procedures of metabolomics include sample collection and pretreatment, metabolomic data collection and acquisition, data processing and analysis, as well as markers identification and biological interpretation. In this context, the review deals with the current development and further trends in the discovery of lung cancer biomarkers based on GC-MS analytical procedures of metabolomics.

## Keywords

Gas Chromatography-Mass Spectrometry (GC-MS), Metabolomics, Lung Cancer Biomarkers

---

# 基于GC-MS的代谢组学方法筛选肺癌标志物的分析流程与研究进展

李彩霞, 张佩佩, 王利超, 刘会君, 潘再法, 王丽丽\*

浙江工业大学化学工程学院, 杭州  
Email: [lili\\_wang@zjut.edu.cn](mailto:lili_wang@zjut.edu.cn)

收稿日期: 2014年3月10日; 修回日期: 2014年3月18日; 录用日期: 2014年3月25日

\*通讯作者。

## 摘要

完整的代谢组学分析流程包括样本的采集与前处理、代谢组数据的采集和获取、数据的处理与分析、标志物的识别与生物学阐释等步骤。本文依据代谢组学的分析流程综述了基于GC-MS的代谢组学方法筛选肺癌标志物的研究流程及进展，并对GC-MS分析技术在代谢组学研究中存在的问题进行了总结与展望。

## 关键词

气相色谱 - 质谱联用, 代谢组学, 肺癌标志物

## 1. 引言

肺癌是对人类健康构成严重威胁的十大恶性肿瘤之一，近年来其发病率和死亡率在我国多数城市位居肿瘤之首[1]。目前，临床上常用的肺癌早期诊断方法主要包括影像学、内镜和分子生物学技术，但這些方法都存在一定的局限性，如灵敏度不高、对人体有害或价格昂贵等[2]。代谢组学是一门依靠现代分析技术，定性定量研究某一生物体系(细胞、组织、体液或生物体)中的代谢产物的新兴学科，其揭示的小分子代谢产物的变化是基体内基因、蛋白质/酶等功能变化的一系列事件的最终结果，直接反映了生物体系的最终状态，从而可以反映机体在特定病理生理状态下整体代谢物质的变化，为疾病的诊断提供了新的研究思路[3]。代谢组学从机体的动态代谢途径中寻找肿瘤标志物，有助于肿瘤的早期发现，目前应用代谢组学方法寻找肺癌标志物的研究层出不穷。研究表明，挥发性代谢产物是肺癌潜在的生物标志物，对肺癌的早期诊断具有重要的意义[4]-[6]。气相色谱-质谱联用技术(GC-MS)作为定性定量检测挥发性化合物最主要的分析手段，被广泛用于复杂生物样本中代谢产物的测定和肿瘤标志物的寻找[7]。

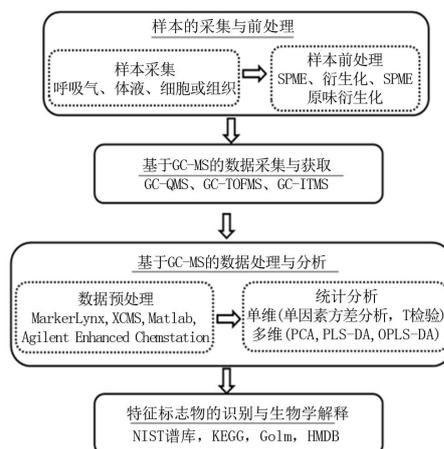
完整的代谢组学研究主要包括样本的采集与前处理、代谢组数据的采集和获取、数据的处理与分析、标志物的识别与生物学阐释等步骤。其研究平台主要由分析技术平台和数据处理平台构成。目前，先进的分析仪器和大量的基础研究使代谢组学分析技术的发展日趋成熟，相比之下，数据处理平台的发展稍显滞后。就基于 GC-MS 的代谢组学方法筛选肿瘤标志物的研究而言，多数研究侧重于分析方法的建立，在数据处理方面多采用单一的单维统计分析找出具有统计学差异的特征代谢物，这在一定程度上缺乏了研究的完整性。近几年，随着化学计量学结合软件技术的发展，先进的数据处理软件和多种统计学分析方法也逐渐被用于基于 GC-MS 的代谢组学方法寻找肿瘤标志物的研究中。为此，本文介绍了基于 GC-MS 的代谢组学方法寻找肿瘤标志物完整的分析流程(图 1)，并围绕该流程综述了基于 GC-MS 的代谢组学方法筛选肺癌标志物的研究过程及进展，同时还探讨了 GC-MS 分析技术在代谢组学研究中存在的问题及展望。

## 2. 分析流程及研究进展

### 2.1. 样本的采集与前处理

#### 2.1.1. 样本的采集

样本的采集和前处理是代谢组学研究中的初始步骤也是最重要的步骤之一。由于代谢组学的研究对象是复杂的生物样本，为了全面详尽的反映生物学过程，代谢组学研究要求严格的实验设计和合适的分析精度。首先，需要准确采集到具有代表性且数量足够的样本，以尽量减少样本个体差异对分析结果的影响。其次，实验设计中对本样本收集的时间、部位、种类、样本群体及样本的保存条件和保存时间等因



**Figure 1.** The workflow of discovering tumor markers by GC-MS-based metabolomic method

**图 1.** 基于 GC-MS 的代谢组学方法寻找肿瘤标志物的分析流程图

素应给予充分考虑。同时，还必需科学、严格、全面地设计和采集对照组样本。另外，在采样过程中，为了避免外界因素引入的系统误差，每次采集样本所用的实验装置要尽量保持干净且一致，采集样本的方法要相同，甚至采样的实验者也最好是同样技术水平或同一工作人员。最后，生物样本采集后，还需要立即进行生物反应灭活处理，即对所收集样本进行快速淬灭，如在液氮或 $-80^{\circ}\text{C}$ 下冷冻、酸碱处理或加有机试剂等处理[8] [9]。

人体呼吸气作为一种直接含有大量挥发性代谢产物的生物来源，很早就被作为基于 GC-MS 代谢组学方法筛选肺癌标志物的研究对象。早在 1985 年，美国的 Gordon 等[10]就采用一套特殊的气体收集装置和计算机辅助的气相色谱仪/质谱分光仪，分析了肺癌患者和对照组呼吸气样品，首次提出并证实了采用呼吸气检测应用于诊断肺癌的思想。随后，O'Neill 等[11]采用聚四氟乙烯样品袋、Phillips[12]用一次性喉舌的便携式呼吸收集装置、以及国内浙江大学的王平课题组[13] [14]、胡燕婕[15]采用 Tedlar 气体采样袋收集人体呼吸气样品，结合 GC-MS 分析，寻找到一些特征的挥发性代谢产物有望作为肺癌标志物。近年来，Santonico 等[16]不仅利用 Tedlar 采样袋收集人体呼吸气，还开发了一种新的支气管镜气体采样装置，成功用于基于 GC-MS 分析人体呼吸气中肺癌标志物的样品采集。

各种易于获取的人体体液，如尿液[17]-[19]、血液[20] [21]及唾液[22]也成为基于 GC-MS 的代谢组学方法寻找肺癌标志物的主要生物样本。此外，胸腔积液作为与肺部疾病密切相关的一种特异的人体体液，近几年也成为基于代谢组学方法筛选肺癌标志物的研究对象[23] [24]。随着研究的深入，其它更复杂的生物样本，如肺癌的肿瘤细胞或组织也逐渐成为基于 GC-MS 的代谢组学研究的样本[25]-[27]。生物体液易于获取，对于疾病的大规模筛查较为便利，病理组织样本的加入，也使得代谢组学在疾病机制机理研究方面有了新的推动力量。但由于生物样本中代谢产物成分复杂且浓度低，痕量的生物标志物往往存在于复杂的背景中，从而增加了基于 GC-MS 代谢组学方法分析肿瘤标志物的样本前处理难度。

### 2.1.2. 样本的前处理

一般需要根据实验对象、实验目的和所采用的仪器分析方法选择合适的样本前处理技术，它直接影响到样本中代谢物的信息能否被准确检出。生物样本的前处理主要包括样品的提取、净化、浓缩等过程。通常采用水或有机溶剂(如甲醇、己烷等)对样品中的代谢产物进行提取，对血液样品还需要通过有机溶剂

沉淀去除样品中所含的蛋白质成分[28], 尿液样品需加入尿素酶去除尿素[29]等操作后再作进一步处理。

GC-MS 分析前, 常采用固相萃取、固相微萃取等前处理技术对样本中微量的挥发性代谢物进行浓缩富集。固相微萃取(SPME)因其集萃取、富集及进样于一体, 成为基于 GC-MS 分析肺癌标志物的代谢组学研究中应用最广的一种样品前处理技术。Wang 等[25]采用 SPME (100  $\mu\text{m}$  PDMS 萃取纤维头) 萃取肺癌患者呼吸气、肿瘤细胞和组织中的挥发性代谢物, 结合 GC-MS 分析, 成功找出 23 种具有潜在诊断价值的肺癌标志物。Ulanowska 等[30]采用 SPME 联合 GC-MS 分析了肺癌患者在接受治疗前后的呼吸气成分, 重点考察了 8 种萃取纤维头对呼吸气样品的萃取效率, 最终选择 CAR/PDMS 萃取纤维头对呼吸气中的挥发性代谢物进行萃取。Rocha 等[31]选用 DVB/CAR/PDMS 的 SPME 纤维头萃取人体尿液中的挥发性代谢物, 经 GC $\times$ GC-TOF MS 分析, 发现一些特征代谢物有望作为生物标志物用于临床试验。最近, Liu 等[24]采用 SPME (75  $\mu\text{m}$  CAR/PDMS 萃取纤维头) 萃取肺癌患者和肺良性疾病患者胸腔积液中的挥发性代谢物, 通过 GC-MS 分析, 寻找到 5 种在肺癌组和肺良性疾病组差异表达的特征代谢物。实验表明, 不同材料、不同膜厚及不同长度的 SPME 纤维头, 可以萃取到的物质种类及含量都会有所不同, 但 SPME 与 GC-MS 联用, 大多只适用于分析生物样本中具有挥发性的代谢物。

由于体液、细胞或组织等生物样本中还存在着大量极性高、挥发性低的代谢物, 如氨基酸、胺、有机酸和芳香族化合物等半挥发性代谢物, 不能直接通过 SPME 联合 GC-MS 分析, 而通过化学衍生可将这些物质转化成极性低、适合 GC-MS 分析的挥发性衍生物。因此, 化学衍生法也是基于 GC-MS 代谢组学方法分析肺癌标志物重要的前处理手段。衍生化处理中通用的是两步法(脲化和衍生化)[32], 常用的衍生化试剂分为硅烷化、酰化和烷基化三类, 基于 GC-MS 代谢组学研究中应用最广泛的是硅烷化试剂。Wedge 等[33]在基于 GC/TOF-MS 分析小细胞癌患者血清和血浆代谢产物的研究中, 首先加入甲醇沉淀去除样品中的蛋白质成分, 再选用 N-甲基三甲基硅基三氟乙酰胺(MSTFA)作为硅烷化试剂对样品中的代谢物进行衍生化处理, 后经离心分离取上层清液用于 GC-MS 分析。Wen 等[34]也采用同样的衍生化前处理操作, 联合 GC-MS 分析了人体血浆中的代谢产物, 发现 37 种在肺腺癌患者和正常健康人血浆中差异表达的代谢物。牛艳洁等[29]选用 N,O-双三甲基硅基三氟乙酰胺(BSTFA)和三甲基硅烷(TMCS)对人体血液和尿液中的小分子代谢物进行硅烷化处理, 结合 GC-MS 分析, 初步研究了初治肺癌与其他肺部疾病患者血清和尿液中的小分子代谢产物, 寻找到血液相关的差异代谢产物 13 种, 尿液相关的差异代谢产物 7 种。最近, Borges 等[35]在利用 GC-MS 分析肺癌患者血浆样品中的多聚糖类代谢物时, 采用二甲亚砜(DMSO)和碘甲烷对样品进行甲基化, 并选用三氟乙酸酐(TFA)对其进行乙酰化处理, 再经液-液萃取、提纯等操作后通过 GC-MS 分析, 成功发现与肺癌相关的聚糖类标志物。

此外, 化学衍生化与 SPME 联用的前处理技术在基于 GC-MS 分析肺癌标志物的代谢组学研究中也有应用。Poli 等[36]采用原位衍生化-固相微萃取的前处理方法, 先将衍生化试剂五氟苯基羟胺(PFBHA)富集到 SPME 萃取头上, 然后将其置于待分析的呼吸气样品中, 在萃取相上同时完成衍生和萃取过程, 再通过 GC-MS 分析了非小细胞癌患者和正常健康人呼吸气中含 3~9 个 C 的直链醛类化合物, 结果发现这些醛类化合物在非小细胞癌患者呼吸气中的浓度均高于正常健康人。Deng 等[21]采用同样的原位衍生化-固相微萃取前处理操作, 结合 GC-MS 定量分析了肺癌患者和正常健康人血液中的己醛和庚醛的含量, 发现二者在肺癌患者血液中的浓度均高于正常健康人。

## 2.2. GC-MS 代谢组数据的采集和获取

完成样本的采集和前处理后, 需借助 GC-MS 分析仪器, 利用 GC 的高分离度和 MS 强大的定性能力, 对样品中挥发性或半挥发性的代谢产物的种类、含量及其变化进行准确的定性和定量分析, 进而得到代谢物谱或代谢指纹图谱。就色谱分析技术而言, 基于 GC-MS 分析肺癌标志物的研究常选用的 GC 毛细管

柱多为非极性柱如 RTX-1 柱[25], 弱极性柱如 DB-5 柱[27] [35]、HP-5 柱[21] [34] [36], 和中等极性柱如 DB-17 柱[37]等。根据质谱技术, GC-MS 通常有气相色谱 - 四级杆质谱(GC-Q MS)、气相色谱 - 飞行时间质谱(GC-TOF MS)和气相色谱 - 离子阱质谱(GC-IT MS)。Q MS 扫描方式有全扫描和选择离子扫描(SIM)之分, 代谢组学研究中对于成分未知的复杂生物样本, 一般采用全扫描模式对样本中指定质量范围内的离子进行全扫描并记录, 得到含有丰富未知代谢物的分子质量和结构信息的质谱图。而在确定了样本中特征代谢物的前提下多采用 SIM 方式对选定的离子进行检测, 以消除样品中其他组分造成的干扰, 提高对特征代谢物的检测灵敏度和定量分析的准确性。TOF MS 提供了更快的扫描率和更高的灵敏度, 采集到的每一个数据点都对应一个完整的质谱图, 检测挥发性化合物的能力比 Q MS 更强, 目前 TOF MS 在代谢组学的研究中应用最为普遍[7]。IT MS 结构小巧, 可以在极低压强下长时间储存离子, 还可以选择任一质量离子进行碰撞解离, 实现二级或多级质谱( $MS^n$ )分析, 具有更强的质谱定性功能。目前, 这三种 GC-MS 联用仪都已被用于肺癌标志物的研究中。

Deng 等[20]利用 GC-Q MS 采用全扫描模式分析了人体血液中的挥发性物质, 发现肺癌病人血液中的己醛和庚醛的浓度远高于正常人。随后, 该课题组还利用 SIM 模式定量分析了血液中的己醛、庚醛, 测定结果重现性 89%~95%, 相对标准偏差(RSD)小于 8.5%, 己醛和庚醛的检测限分别达  $0.006 \text{ nml}\cdot\text{L}^{-1}$ 、 $0.005 \text{ nml}\cdot\text{L}^{-1}$ [21]。最近, Hecht 等[37]采用气相色谱 - 负离子化学电离 - 串联四极杆质谱(GC-NICI-MS/MS), 分析了致癌多环芳烃菲[D10]在吸烟者(肺癌高危人群)尿液中的代谢产物 PheT 的浓度水平, 以评价吸烟人群对菲[D10]的代谢能力。Buszewski 等[38]采用 GC-TOF MS 定性定量分析了肺癌患者和正常人体呼吸气中痕量的挥发性代谢产物, 该方法可快速检出人体呼吸气中 ppb 级别含量的挥发性代谢产物, 线性范围 0.83~234.05 ppb, 检测限 0.31~0.75 ppb, RSD 小于 10.0%, 结果发现乙酸乙酯和 2-戊酮与肺癌的发生有关。Kischkel 等[39]采用 GC-IT MS 分析了肺癌患者在单肺通气治疗期间的呼吸气成分, 结果发现丙酮、丁烷和戊烷等挥发性代谢产物可作为内源性的肺癌标志物。

### 2.3. GC-MS 代谢谱数据的处理与分析

代谢组学的数据处理与分析是指使用数据提取软件结合化学计量学工具对仪器分析所获得的代谢组数据进行信息挖掘和统计分析。经过 GC-MS 仪器分析, 样本的代谢物信息由仪器产生的谱图信号呈现出来。基于 MS 平台的仪器所获得的生物样本谱图通常含有几百到几千个色谱峰, 且每个色谱峰至少有 3~5 次扫描, 甚至有些精密仪器扫描次数达 20~50 次, 而质谱峰间隔也有若干次扫描。因此, 一张色谱图可能含有成千上万张质谱图。并且实际的代谢组学研究中一个实验通常需要完成批量样品的平行测定, 再将这些样品的仪器平台信息进行整合处理, 那么工作量相当庞大。所以, 由 GC-MS 仪器分析得到的代谢谱数据具有多维复杂性, 需要选择合适的数据处理方法从中挖掘出有效的信息。

#### 2.3.1. GC-MS 代谢谱数据预处理

数据预处理主要是指数据的提取, 即将高灵敏度的仪器采集到的图谱信号, 经过软件处理转变成为可计算的数据。基于 GC-MS 分析产生的代谢谱数据是三维的, 分别是保留时间、质荷比( $m/z$ )和峰强度。这类数据不能直接用于统计分析, 还需要进行一系列的预处理, 包括滤噪、峰检测, 重叠峰解析, 峰对齐、匹配, 归一化和标准化等操作[40], 这些过程常借助一系列数据预处理软件来完成。目前, 就色-质联用的数据预处理而言, 相应的软件多达几十种, 主要可分为两大类: 一类是仪器供应商开发的软件, 如 Waters MarkerLynx、Agilent Enhanced Chemstation 以及 AB MarkerView 等; 另一类是独立于仪器的软件, 如 MetAlign、XCMS Online 以及 MetaboAnalyst 等。其中, 前者属于商业软件, 功能强大, 通常能为代谢组学研究提供从样品分析到数据处理的整套方案, 但对数据格式要求严格且价格昂贵, 一般不能处理其它品牌仪器产生的数据; 后者则是用于提取数据的免费软件, 但在使用这些软件时, 使用者需要

一定的软件基础知识和较好的英文识别能力, 通常需要根据要求将仪器所产生的图谱信号转换成可供读取的数据格式。此外, 有研究者还根据实验数据处理的需求自主开发设计了相应的编程软件, 如美国 Math Works 公司编写的 Matlab 软件[41], 浙江大学程翼宇研究组自主编写的用于代谢组学研究的相关程序[42]。

目前, 这些软件逐渐被应用于基于色-质联用技术的生物标志物的代谢组学研究中。赵丽丽等[43]采用 MarkerLynx 软件处理基于超高效液相色谱-质谱采集的肺癌患者血浆的代谢谱数据, 对原始数据进行了峰匹配和检测、及归一化处理, 结合统计学分析方法, 筛选出潜在的肺癌标志物。近几年, 在处理基于 GC-MS 的代谢组学方法寻找肿瘤标志物的研究中, XCMS 软件也逐渐受到研究者的青睐。XCMS 软件是由美国 Scripps 研究所开发的现已发展成网上在线分析平台 XCMS Online (<https://xcmsonline.scripps.edu/index.php>)的一种数据处理软件, 是一种基于 R 语言的全英文环境的质谱数据预处理工具, 可处理多家公司生产的 GC-MS 和 LC-MS 仪器产生的图谱数据[44]。XCMS 软件支持多参数设置, 可以通过对参数的设置完成代谢谱中各物质保留时间的校正、峰信息识别及提取和初步的统计分析等操作, 同时, XCMS 软件还具有强大的数据可视化功能。最近, Abaffy 等[45]在基于 GC-MS 代谢组学方法分析黑色素瘤皮肤与健康皮肤特征代谢物的初步研究中, 借助 XCMS Online 对 GC-MS 仪器分析产生的代谢谱数据进行处理, 发现月桂酸和棕榈酸在黑色素瘤皮肤中具有较高的水平。此外, 在分析尿液中挥发性肺癌标志物的研究中, Hanai 等[19]采用 XCMS 软件对基于 GC-TOF MS 的代谢谱数据进行了处理, 从中分离出 19 组具有统计学差异的特征代谢物信息。同时, 该课题组还利用 GC-MS 分析了人体肺癌细胞和植入肺癌肿瘤的小鼠尿液中的挥发性代谢物, 结合 XCMS 软件进行数据处理, 发现丁二酸二甲酯、2-戊酮等 7 种潜在的肺癌标志物[26]。

### 2.3.2. 数据统计分析

预处理后的数据可供后续单维或多维统计分析。单维统计方法用于分析每个变量在整批样本中的特征, 如某变量在疾病组和对对照组的均值差异等。而多维统计方法用于分析整个数据集, 一是建立分类模型, 对不同性质的样本(正常/患病)进行正确的区分; 二是进行特征提取, 评价每个变量每个样本的重要性, 关注变量之间的关系, 最终找到对于分类贡献显著的差异变量。常用的单维统计方法包括相关分析、受试者工作特征曲线分析(ROC)以及各种参数和非参数检验, 如单因素方差分析、T 检验和曼-惠特尼检验等。但由于代谢组学研究所采集的数据往往数据量大且维数多, 并且代谢组学研究的多是批量样本的分析, 因此传统的单维统计分析往往难以胜任, 为了挖掘所获数据的本质和潜在信息, 常常结合多维统计分析进行数据处理[46]。代谢组学研究中常用的多维统计方法主要是模式识别技术, 包括无监督模式识别和有监督模式识别。

主成分分析(principle component analysis, PCA)是代谢组学领域常用也是比较有效的一种无监督模式识别方法, 它是一种基于投影技术对数据进行降维处理的分析方法。经过 PCA 分析, 得到的主成分是由具有相关性的原始变量通过线性组合而成的相互正交的新变量, 一般只需要选取前面的几个主成分, 使其能提供原始数据 80% 以上的信息即可。各样本在各个主成分的得分就是其在计算的数学模型中的空间坐标, 直观地反映了各个样品在数学模型空间中的分布情况。无监督模式识别是指在不对样本加以分组的情况下进行数据分析, 有利于了解数据的整体情况并对数据从整体上进行把握, 特别有利于发现和剔除异常样品, 提高模型的准确性[47]。而有监督模式识别就是先将检测样品按照类别进行分组, 再进行分析, 在计算数学模型时计算机就把各组加以区分, 此法的特点是忽略组内的随机差异, 突出样本组间差异。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)都是代谢组学中典型的有监督模式识别方法。数据处理时需要根据实验样本的特点选择合适的模式识别技术: 当样本的组间差异较大, 而组内差异较小时, 无监督模式识别分析可以明显区分组间差异; 但当样品的组间差异不明显, 而组内差异较大时, 有监督模式识别分析更易于发现和区分组间差异[47]。目前, 应用的较广的模式识别

软件为 SIMCA-P+, 它对多种软件导出的数据都有较好的兼容性。Chen 等[48]利用 SIMCA-P 11.0 软件处理基于 LC-MS 分析的人体尿液代谢组学数据时, 先后采用了 PCA 和 PLS-DA 两种方法对肝癌组和健康组的尿液代谢谱进行模式识别分析, 结果显示 PCA 得分图无法将两样本组区分开, 而 PLS-DA 却可以很好的将两组样本区分开, 表明有监督的模式识别方法在一定程度上提高了模型的分类效能。Song 等[49]在用 GC-MS 对胃癌患者和健康人血清代谢组学的分析中, 采用 OPLS-DA 在 SIMCA-P 11.5 软件中实现对代谢组学数据的多维统计分析, 结合 VIP > 1 原则或 Wilcoxon 检验, 找出两组间的差异代谢物。在基于 GC-MS 的代谢组学方法分析肺癌标志物的研究中, 牛艳洁等[29]利用 SIMCA-P 12.0+ 软件采用 OPLS-DA 进行建模分析, 结合 T 检验比较肺癌组和其它肺部疾病组的代谢产物水平, 成功找出血清、尿液相关的差异代谢物。

#### 2.4. 特征代谢物的鉴定及生物学阐释

经过仪器分析和一系列的数据处理, 发现实验样本间的差异并找出引起这种差异的特征代谢物后, 需要对这些特征代谢物进行化学结构鉴定, 并通过查阅生物学知识和相关功能数据库, 推断该代谢物的生物功能、代谢节点和相关代谢途径等信息。基于质谱技术的代谢组学研究, 很大的一个优势就是可以将通过代谢物质谱信号的化学鉴定得到的数据转化为生物代谢途径信息, 但面临的一个大的难题也是对谱学信号中代谢物的分子鉴定。在 GC-MS 分析中, 电子轰击电离(EI)是最常用的电离技术, 化合物在 EI 作用下可产生特征性的质谱裂解方式, 目前已经有成熟的经典 EI 谱库可供检索, 如 NIST 谱库。因此, 对基于 GC-MS 的代谢组学研究而言, 代谢物的物质鉴定主要基于 NIST 质谱数据库, 通过谱库检索功能初步鉴定代谢物的分子结构, 再利用与之相对应的可靠标准品进一步验证。目前, 除了仪器自带的质谱数据库外, 国际上还有一些专用的生化数据库[50], 可用于未知代谢物的结构鉴定和已知代谢物的生物功能解释, 如京都基因与基因组百科全书 KEGG、Golm 和 HMDB 等。此外, 我国也有一些研究者致力于开发专门的 GC-MS 代谢物质谱数据库, 如中南大学化学化工学院中药现代化研究中心参考 David S. Wishart 等建立的人类代谢组学数据库([www.hmdb.ca](http://www.hmdb.ca)), 建立了人类代谢组学 GC-EIMS 标准物质数据库[51]。

### 3. 结论与展望

综上所述, 基于 GC-MS 联用技术的代谢组学研究主要由 GC-MS 分析技术平台和数据处理平台构成。在分析技术方面, 由于 GC-MS 主要分析的是具有挥发性或经化学衍生具有挥发性的化合物, 而生物样本中大部分代谢产物是不易挥发的, 且多数代谢产物微量存在于复杂的背景中, 因此需要繁复的提取、净化、浓缩和衍生化等前处理操作。此外, 利用 GC-MS 分析成份复杂的生物样本时, GC 会出现无法完全分离各种物质的现象。对此, 需要进一步深入研究, 如开发新的衍生化试剂, 使其能同时与多个官能团发生衍生化反应, 或者开发新型的 SPME 萃取头, 使其集衍生化和萃取于一体, 或者不经衍生化就能具有更高的萃取效率和更广的适用范围, 以获得操作简单、重现性更好的前处理方法; 或者开发新型高效的专用色谱柱, 使其适合于大量复杂生物样本的分离与分析, 以改善实验结果的可靠性。另外, 如何获得可参照的内标物实现所有代谢物的绝对定量分析, 也将成为 GC-MS 应用于代谢组学中的研究重点。在数据处理方面, 虽然化学计量学方法结合软件技术的开发在近几年得到了快速发展, 但在实际研究中的应用还是稍显滞后。这或许是由于完整的代谢组学数据处理平台对跨学科知识的综合要求, 研究者不仅要具备一定的分析化学知识, 还需要相关的化学计量学、医学生物学甚至计算机科学等多方面的知识。所以, 组织并开设专门的代谢组学数据处理专题培训也是必要的。另外, 开发出更先进的数据处理软件和可用于代谢物定性和生物学信息参考的专用数据库仍是今后代谢组学研究的主要方向。总之, 基于 GC-MS 的代谢组学研究需要更多更好的分离分析手段, 更有力的成分分析仪器, 更强大的数据挖掘和处理工具, 以及更完备的代谢组学数据库。

## 项目基金

浙江省分析测试基金(No. 2011C37009); 浙江省重中之重学科开放基金(No. 200908)。

## 参考文献 (References)

- [1] Chen, W.Q., Zheng, R.S., Zhang, S.W., et al. (2013) Report of incidence and mortality in China cancer registries, 2009. *Chinese Journal of Cancer Research*, **25**, 10-21.
- [2] 胡海洋, 林之峰, 林强 (2013) 肺癌早期诊断的研究进展. *现代生物医学进展*, **14**, 2768-2773.
- [3] Nicholson, J.K., Connelly, J., Lindon, J.C., et al. (2002) Metabonomics: A platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery*, **1**, 153-161.
- [4] Miekisch, W., Schubert, J.K. and Noeldgeschomburg, G.F. (2004) Diagnostic potential of breath analysis-focus on volatile organic compounds. *Clinica Chimica Acta*, **347**, 25-39.
- [5] Buszewski, B., Keszy, M., Ligor, T., et al. (2007) Human exhaled air analytics: Biomarkers of diseases. *Biomedical Chromatography*, **21**, 553-566.
- [6] Silva, C.L., Passos, M. and Camara, J.S. (2011) Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry. *British Journal of Cancer*, **105**, 1894-1904.
- [7] 李娟, 任路静, 孙冠男, 等 (2013) 气相色谱 - 质谱联用技术及其在代谢组学中的应用. *生物工程学报*, **4**, 434-446.
- [8] Bando, K., Kawahara, R., Kunimatsu, T., et al. (2010) Influences of biofluid sample collection and handling procedures on GC-MS based metabolomic studies. *Journal of Bioscience and Bioengineering*, **110**, 491-499.
- [9] 汪明明, 程海婷, 薛明 (2011) 基于 LC-MS 的代谢组学分析流程与技术方法. *国际药学研究杂志*, **2**, 130-136.
- [10] Gordon, S.M., Szidon, J.P., Krotoszynski, B.K., et al. (1985) Volatile organic compounds in exhaled air from patients with lung cancer. *Clinical Chemistry*, **31**, 1278-1282.
- [11] O'neill, H.J., Gordon, S.M., O'neill, M.H., et al. (1988) A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. *Clinical Chemistry*, **34**, 1613-1618.
- [12] Phillips, M., Altorki, N., Austin, J.H., et al. (2007) Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomarkers*, **3**, 95-109.
- [13] 徐凤娟 (2008) 肺部呼吸诊断特征气体的研究. 硕士论文, 浙江大学, 杭州, 47-66.
- [14] 於锦 (2011) 呼出气体及其冷凝物中肺癌标志物及其检测方法的研究. 硕士论文, 浙江大学, 杭州, 23-44.
- [15] 胡燕婕 (2010) 肺癌呼气特征性 VOCs 的筛选及诊断价值研究. 博士论文, 浙江大学, 杭州, 1-52.
- [16] Santonico, M., Lucantoni, G., Pennazza, G., et al. (2012) *In Situ* detection of lung cancer volatile fingerprints using bronchoscopic air-sampling. *Lung Cancer*, **77**, 46-50.
- [17] Matsumura, K., Opiekun, M., Oka, H., et al. (2010) Urinary volatile compounds as biomarkers for lung cancer: A proof of principle study using odor signatures in mouse models of lung cancer. *PLoS ONE*, **5**, 1-11.
- [18] Guadagni, R., Miraglia, N., Simonelli, A., et al. (2011) Solid-phase microextraction-gas chromatography-mass spectrometry method validation for the determination of endogenous substances: Urinary hexanal and heptanal as lung tumor biomarkers. *Analytica Chimica Acta*, **701**, 29-36.
- [19] Hanai, Y., Shimono, K., Matsumura, K., et al. (2012) Urinary volatile compound as biomarkers for lung cancer. *Bio-science, Biotechnology, and Biochemistry*, **76**, 679-684.
- [20] Deng, C.H., Zhang, X.M. and Li, N. (2004) Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography-mass spectrometry. *Journal of Chromatography B*, **808**, 269-277.
- [21] Deng, C.H., Li, N. and Zhang, X.M. (2004) Development of headspace solid-phase microextraction with on-fiber derivatization for determination of hexanal and heptanal in human blood. *Journal of Chromatography B*, **813**, 47-52.
- [22] Sanchez, M.D., Garcia, E.H., Pavon, J.L.P., et al. (2012) Fast analytical methodology based on mass spectrometry for the determination of volatile biomarkers in saliva. *Analytical Chemistry*, **84**, 379-385.
- [23] Zhou, X.M., He, C.C., Liu, Y.M., et al. (2012) Metabonomic classification and detection of small molecule biomarkers of malignant pleural effusions. *Analytical and Bioanalytical Chemistry*, **404**, 3123-3133.
- [24] Liu, H.J., Wang, H., Li, C.X., et al. (2014) Investigation of volatile organic metabolites in lung cancer pleural effusions by solid-phase microextraction and gas chromatography/mass spectrometry. *Journal of Chromatography B*, **945-946**, 53-59.

- [25] Wang, Y.S., Hu, Y.J., Wang, D., et al. (2012) The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomarkers*, **11**, 129-137.
- [26] Hanai, Y., Shimono, K., Oka, H., et al. (2012) Analysis of volatile organic compounds released from human lung cancer cells and from the urine of tumor-bearing mice. *Cancer Cell International*, **12**, 7.
- [27] Naz, S., Garcia, A. and Barbas, C. (2013) Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Analytical Chemistry*, **85**, 10941-10948.
- [28] 蔡品品, 卢红梅 (2013) 气相色谱-质谱分析血清代谢组前处理中的除蛋白及衍生方法. *分析化学*, **8**, 1183-1187.
- [29] 牛艳洁, 江银玲, 许长江, 等 (2012) 代谢组学方法分析肺癌患者血清和尿液小分子代谢产物的初步研究. *中国肺癌杂志*, **4**, 195-201.
- [30] Ulanowska, A., Trawinska, E., Sawrycki, P., et al. (2012) Chemotherapy control by breath profile with application of SPME-GC/MS method. *Journal of Separation Science*, **35**, 2908-2913.
- [31] Rocha, S.M., Caldeira, M., Carrola, J., et al. (2012) Exploring the human urine metabolomic potentialities by comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry. *Journal of Chromatography A*, **1252**, 155-163.
- [32] Dunn, W.B., Broadhurst, D., Begley, P., et al. (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, **6**, 1060-1083.
- [33] Wedge, D.C., Allwood, J.W., Dunn, W., et al. (2011) Is serum or plasma more appropriate for intersubject comparisons in metabolomic studies? An assessment in patients with small-cell. *Analytical Chemistry*, **83**, 6689-6697.
- [34] Wen, T., Gao, L., Wen, Z.M., et al. (2013) Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. *Molecular BioSystems*, **9**, 2370-2378.
- [35] Borges, C.R., Rehder, D.S. and Boffetta, P. (2013) Multiplexed surrogate analysis of glycotransferase activity in whole biospecimens. *Analytical Chemistry*, **85**, 2927-2936.
- [36] Poli, D., Goldoni, M., Corradi, M., et al. (2010) Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatization SPME-GC/MS. *Journal of Chromatography B*, **878**, 2643-2651.
- [37] Hecht, S.S., Hochalter, J.B., Carmella, S.G., et al. (2013) Longitudinal study of [D<sub>10</sub>] phenanthrene metabolism by the diol epoxide pathway in smokers. *Biomarkers*, **18**, 144-150.
- [38] Buszewski, B., Ligor, T., Jezierski, T., et al. (2012) Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: Comparison with discrimination by canines. *Analytical and Bioanalytical Chemistry*, **404**, 141-146.
- [39] Kischkel, S., Miekisch, W., Fuchs, P., et al. (2012) Breath analysis during one-lung ventilation in cancer patients. *European Respiratory Journal*, **40**, 706-713.
- [40] 许国旺, 等 (2008) 代谢组学—方法与应用. 科学出版社, 北京, 38-40.
- [41] Kleckner, I.R. and Foster, M.P. (2012) GUARDD: User-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. *Journal of Biomolecular NMR*, **52**, 11-22.
- [42] 黄欣, 龚益飞, 虞科, 等 (2007) 基于气相色谱-质谱的代谢组学方法研究四氯化碳致小鼠急性肝损伤. *分析化学*, **12**, 1736-1740.
- [43] 赵丽丽 (2009) 代谢组学在肺癌中的临床应用研究. 硕士学位论文, 大连医科大学, 大连, 15-30.
- [44] Tautenhahn, R., Patti, G.J., Rinehart, D., et al. (2012) XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, **84**, 5035-5039.
- [45] Abaffy, T., Moller, M.G., Riemer, D.D., et al. (2013) Comparative analysis of volatile metabolomics signals from melanoma and benign skin: A pilot study. *Metabolomics*, **9**, 998-1008.
- [46] Trygg, J., Holmes, E. and Lundstedt, T. (2007) Chemometrics in metabonomics. *Journal of Proteome Research*, **6**, 469-479.
- [47] 阿基业 (2010) 代谢组学数据处理方法—主成分分析. *中国临床药理学与治疗学*, **5**, 481-489.
- [48] Chen, J., Wang, W.Z., Lv, S., et al. (2009) Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Analytica Chimica Acta*, **650**, 3-9.
- [49] Song, H., Peng, J.S., Yao, D.S., et al. (2012) Serum metabolic profiling of human gastric cancer based on gas chromatography/mass spectrometry. *Brazilian Journal of Medical and Biological Research*, **45**, 78-85.
- [50] GO, E.P. (2010) Database resources in metabolomics: An overview. *Journal of Neuroimmune Pharmacology*, **5**, 18-30.
- [51] 张良晓, 陈爱明, 梁逸曾, 等 (2008) 人类代谢组学 GC-EIMS 标准物质数据库的建立. *第二届全国生命分析化学学术报告与研讨会*, F12, 292.