DOI:10.13602/j.cnki.jcls.2021.12.06

·“迈瑞杯”临检发展专栏·

MALDI-TOF 质谱在乳腺癌诊断及分型中的应用

许小雨[1,2],宋御繁[2],孙克娜[2],陈英剑[2],胡成进[2](1.潍坊医学院医学检验学院,山东潍坊 261053;2.联勤保障部队第九六〇医院检验科,济南 250031)

摘要:目的　应用基质辅助激光解吸电离飞行时间质谱(MALDI-TOF-MS)技术建立乳腺癌诊断预测模型,寻找乳腺癌血清中的差异蛋白质/多肽,并区分三阴性乳腺癌(TNBC)与非三阴性乳腺癌(NTNBC)。**方法**　共纳入 61 例乳腺癌患者、60 例乳腺纤维腺瘤患者及 44 例体检健康者。按照 3∶1 的比例随机分为训练组(乳腺癌患者 45 例,体检健康者 33 例)和验证组(乳腺癌患者 16 例,健康人对照 11 例)。采用弱阳离子交换磁珠(WCX-MB)提取血清蛋白质/肽,基质辅助激光解吸电离飞行时间质谱(MALDI-TOF-MS)技术检测并筛选出差异峰,构建诊断预测模型并初步验证。观察分析乳腺癌患者血清蛋白质谱分子变化。**结果**　应用 ClinProTools 软件构建了乳腺癌组与健康组的诊断预测模型。遗传算法(GA)模型的特异性为 90.9%,敏感性为 93.8%,准确性为 92.6%。对纤维腺瘤组和对照组、乳腺癌组和纤维腺瘤组以及 TNBC 组和 NTNBC 组分别进行聚类分析,发现有显著差异峰可将其准确区分:差异峰 m/z 4 220.9、5 920.36 在纤维腺瘤组中的峰强度明显高于对照组,但其在乳腺癌组中的峰强度明显低于纤维腺瘤组;差异峰 m/z 4 220.9、5 920.36 在 TNBC 组中的峰强度明显高于 NTNBC 组。**结论**　应用 MALDI-TOF-MS 技术建立的乳腺癌诊断预测模型具有高敏感性、高特异性的特点,该技术可用于乳腺癌患者的鉴别诊断以及区分乳腺癌的基因分型。

关键词:基质辅助激光解吸电离飞行时间质谱;诊断预测模型;血清;三阴性乳腺癌;乳腺癌

中图分类号:R737.9;R446.9　　**文献标志码:**A

Application of MALDI-TOF mass spectrometry in diagnosis and classification of breast cancer

XU Xiaoyu[1,2], *SONG Yufan*[2], *SUN Kena*[2], *CHEN Yingjian*[2], *HU Chengjin*[2](1.*Department of Medical Laboratory, Weifang Medical University, Weifang* 261053, *Shandong*; 2. *Clinical laboratory of the* 960*th Joint Logistic Support Force Hospital, Jinan* 250031, *Shandong, China*)

Abstract: Objective　To establish a diagnosis and prediction model of breast cancer by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), to search for differential proteins/peptides in serum of breast cancer, and to distinguish triple negative breast cancer (TNBC) from non-triple negative breast cancer (NTNBC). **Methods**　61 breast cancer patients, 60 breast fibroadenoma patients and 44 healthy subjects were included in the study. The patients were randomly divided into training group (45 cases of breast cancer, 33 cases of healthy subjects) and verification group (16 cases of breast cancer, 11 cases of healthy subjects) according to the ratio of 3∶1. After the serum protein/peptide was extracted by weak cation exchange magnetic beads (WCX-MB), matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was used to detect and screen the difference peaks, and the diagnostic prediction model was established and preliminatively verified. To observe and analyze the molecular changes of serum protein profile in patients with breast cancer. **Results**　ClinProTools provided diagnostic prediction models for breast cancer group and healthy group. The specificity, sensitivity and accuracy of genetic algorithm (GA) model were 90.9%, 93.8% and 92.6% respectively. Cluster analysis was conducted for fibroadenoma group and control group, breast cancer group and fibroadenoma group, and TNBC group and NTNBC group respectively, and significant difference peaks were found to be able to accurately distinguish them: The peak intensity of m/z 4 220.9 and 5 920.36 in fibroadenoma group was significantly higher than that in control group, but the peak intensity of m/z 4 220.9 and 5 920.36 in breast cancer group was significantly lower than that in fibroadenoma group. The peak intensity of m/z 4 220.9 and 5 920.36 in TNBC group was significantly higher than that in NTNBC group. **Conclusion**　The diagnosis and prediction model of breast cancer established by MALDI-TOF-MS has the characteristics of high sensitivity and high specificity, which can be used to distinguish breast cancer patients from healthy people and benign disease groups, and can also be used to distinguish breast cancer genotyping.

Key words: matrix-assisted laser desorption ionization time-of-flight mass spectrometry; diagnostic prediction model; serum; triple-negative breast cancer; breast cancer

作者简介:许小雨,1995 年生,女,硕士研究生,研究方向为重大疾病分子诊断及其机制。

通信作者:胡成进,主任医师,博士,Email:hcj6289@163.com。

乳腺癌是导致世界女性癌症死亡最主要的原因,乳腺癌常用的检查方法有活检、乳房 X 线检查、CT 检查、超声检查和磁共振成像(MRI)等,然而上述方法昂贵、耗时,不适合大规模患者的早期筛查[1]。由美国食品药品监督管理局(Food and Drug Administration, FDA)认证的生物学标志物——CA153 已应用于临床,但其特异性和敏感性仅为 65.7%和 76.7%[2],且其仅可用于监测乳腺癌的疗效、预后评估以及复发判断。目前临床上主要通过免疫组化染色法检测肿瘤细胞雌激素受体(ER),孕酮受体(PR)和人类表皮生长因子受体(HER2)蛋白质的表达,并根据结果将乳腺癌区分为不同的乳腺癌亚型,但此种方法成本高、操作繁琐且需进行染色处理,实际操作中存在很多因素影响免疫组化染色结果的准确性和稳定性。因此,临床上需要寻找到新的方法以提高乳腺癌诊断阳性率,并为临床个体化治疗提供决策依据。

基质辅助激光解吸电离飞行时间质谱(MALDI-TOF-MS)是一种蛋白质组学研究的有效工具,在肿瘤精准诊疗中具有重要作用。本研究旨在采用 MALDI-TOF-MS 技术建立一种具有高敏感性和高特异性,以及适用于临床乳腺癌诊断的预测模型,应用该技术将乳腺癌患者从体检健康者及良性疾病者中鉴别出,并进一步区分三阴性乳腺癌(TNBC)和非三阴性乳腺癌(NTNBC)患者。

1 材料及方法

1.1 一般资料 收集 2020 年 8 月至 2021 年 3 月联勤保障部队某医院经病理组织学活检确诊的乳腺癌患者(61 例)和纤维腺瘤患者(60 例),病理分型为浸润性导管癌 47 例、浸润性小叶癌 1 例、导管内原位癌 8 例及浸润性导管癌伴高级别导管内癌 5 例。乳腺癌分子分型分为 TNBC 10 例, NTNBC 51 例。乳腺癌患者年龄 32~77 岁,中位年龄 53 岁;乳腺纤维腺瘤患者年龄 33~74 岁,中位年龄 55 岁。病例组纳入标准:(1)患者初次经病理组织学活检确诊为乳腺癌或纤维腺瘤;(2)既往未经其他手术、抗癌药物及放化疗治疗;(3)具有完整的临床病理学诊断资料;(4)无其他恶性肿瘤病史。排除标准:(1)非乳腺癌及纤维腺瘤患者;(2)接受过抗癌治疗者;(3)合并其他恶性肿瘤者;(4)感染性疾病患者。以同期 44 名体检健康者作为健康人对照组,年龄 35~79 岁,中位年龄 55 岁。本研究经解放军第九六〇医院医学伦理委员会审核批准[(2021)科研伦理审第(60)号],各研究对象均知情同意。

1.2 仪器和试剂 基质辅助激光解吸电离飞行时间质谱系统(MALDI-TOF,德国 Bruker 公司),低温高速离心机(德国 Eppendorf 公司),电子天平(瑞士 Mettler Toledo 公司)。弱阳离子交换磁珠(WCX-MB)试剂盒、蛋白质/多肽校准品(德国 Bruker 公司),α-氰基-4 羟基肉桂酸(HCCA)、三氟乙酸(TFA)、乙腈(ACN)、无水乙醇(美国 Sigma 公司)。

1.3 方法

1.3.1 标本采集及预处理 在各受试者治疗前采集清晨空腹静脉血 5 mL 贮于含有惰性分离胶的真空采血管中,于室温条件下静置 30 min,置于离心机中以 3 000 r/min 离心 5 min,取血清分装至 0.5 mL的 Ep 管中,随后置于-80 ℃冻存备用。

1.3.2 血清标本分组 将所有的样本按照 3∶1 的比例进行随机分组,训练组 78 例(乳腺癌患者 45 例,体检健康者 33 例),验证组 27 例(乳腺癌患者 16 例,体检健康者 11 例)。训练组用以建立诊断预测模型,验证组用以评估模型诊断效能。

1.3.3 血清预处理 从-80 ℃冰箱取出血清标本,置于-20 ℃冻融 8 h,再置于 4 ℃冻融 6 h,将复融的血清以 10 000 r/min 离心 3 min。进行以下操作:每 5 μL 血清标本与 10 μL 弱阳离子交换磁珠(WCX-MB)和 10 μL 结合缓冲液(BB)混合于 200 μL试管中。混合后,将样品管放入磁力分离器中,1 min 后弃除试管中的上清液。样本与磁珠结合后,用 100 μL 的洗涤缓冲液(WB)洗涤 3 次,加入 5 μL 的磁珠洗脱缓冲液(EB)与磁珠结合。悬浮液样本从上清液中完全分离后转移至新的 10 μL 试管中。每个试管中加入 5 μL 稳定缓冲液(SB),用于质谱分析。

1.3.4 仪器校准与质量控制 检测前使用蛋白质/多肽校准品混合液进行仪器校准,随机选取 30 例体检健康者血清标本,取 5 μL 混匀,用于评价质谱仪的稳定性。弱阳离子磁珠(WCX-MB)提取混合血清中的蛋白质/多肽,检测时每 15 个样本进行 1 次点靶,质谱上样检测获得质谱峰,计算每批质谱多肽峰的变异系数(*CV*)。

1.3.5 血清处理标本质谱检测 取提取的蛋白质/多肽样本 1 μL 点样于 AnchorChip 384 靶板上,样本一式三份点靶,自然干燥后取 1 μL 新鲜配置的 6 mg/mL基质(6 mg HCCA,50%ACN,2%TFA)覆盖于样本点上,自然晾干后置于质谱仪进行检测。在

正离子线性模式下，1 ~ 10 kDa 分子质量范围内采集峰的 m/z 值和强度。采用的离子源 1 为 19.64 kV，离子源 2 为 18.34 Kv，基质抑制为 m/z 700 Da。在激光能量为 75%下采集谱图，每个样本点累积轰击 1 200 次。

1.3.6　峰谱图处理　使用 ClinProTools 软件及 FlexAnalysis软件对采集的原始质谱图进行平滑、去噪、基线去除和归一化处理，并确定分子质量为 1 ~ 10 kDa 的峰，剔除信噪比(S/N)<5 的峰。

1.3.7　建立及验证乳腺癌诊断预测模型　采用 ClinProTools 软件中包含的 3 种模型算法：遗传(GA)算法、监督神经网络(SNN)算法及快速分类(QC)算法对乳腺癌建立诊断预测模型，并使用验证血清组血清标本对模型进行验证，得出模型的敏感性、特异性及准确性，从中选出最佳诊断预测模型。

2　结果

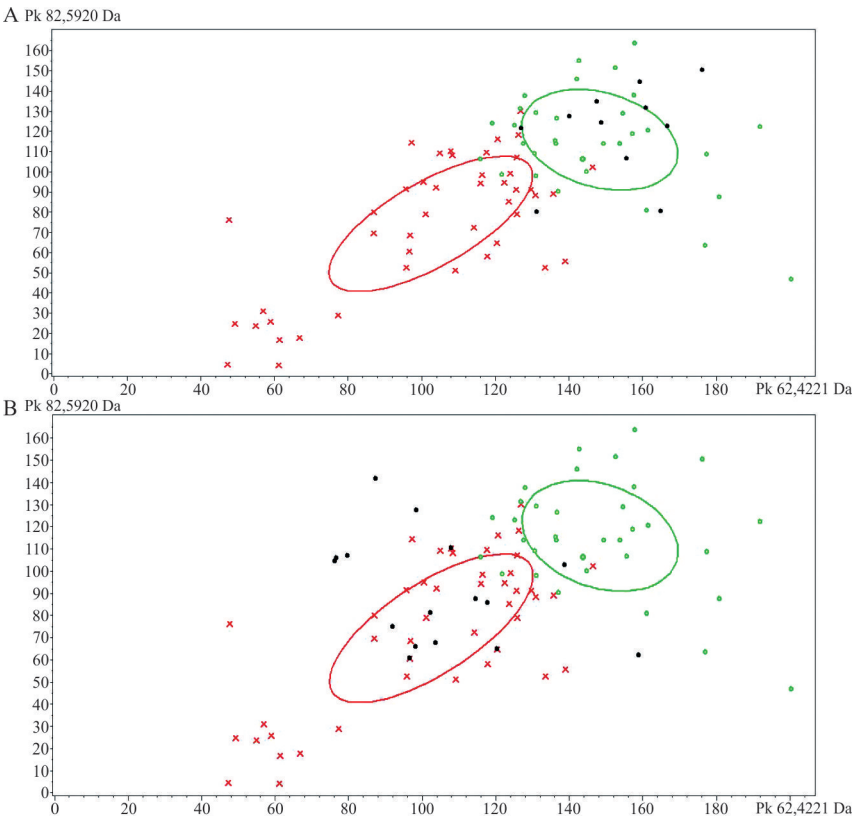
2.1　质量控制结果分析　本次实验检测时每 15 个样本点 1 次靶，检测后得到 9 个质控样本质谱峰，计算 1 ~ 10 kDa 范围内 9 个峰的平均 *CV* 为 19.56%(11.67% ~ 29.78%)。表明此 *CV* 在正常范围内，此系统检测一致性较好，符合要求。

2.2　乳腺癌诊断预测模型的建立与验证结果分析　应用 ClinProTools 软件对乳腺癌组与对照组进行分析，建立了诊断预测模型，并使用验证组进行验证(表 1)。结果发现 GA 算法建立的模型诊断效能最佳。选取差异最显著的 2 个峰：m/z 5 920.36、m/z 4 220.9 分别为 *X*、*Y* 轴进行聚类分析，结果显示乳腺癌组与对照组可明显区分(图 1)，在 GA 算法模型下使用验证组血清验证，16 例乳腺癌患者中正确判断 15 例，误判 1 例；11 例对照者中正确判断 10 例，误判 1 例。GA 模型的预测诊断特异性为 90.9%，敏感性为 93.8%，准确性为 92.6%。

表 1　3 种诊断预测模型的效能(%)

运算方法	识别能力	交叉验证	敏感性	特异性	准确性
GA	98.9	91.1	93.8	90.9	92.6
SNN	97.4	86.1	81.3	90.9	85.2
QC	84.4	83.3	50.0	81.8	63.0

注：GA 为遗传算法；SNN 为监督神经网络算法；QC 为快速分类算法。



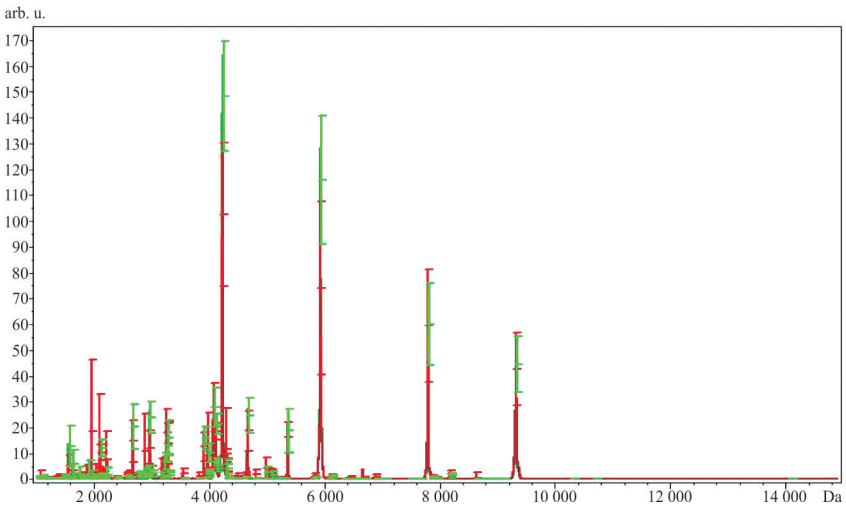
注：红色叉号为训练组乳腺癌患者；绿色圆圈为训练组对照者，A 中黑点为验证组乳腺癌患者；B 中黑点为验证组对照者。

图 1　训练组样本聚类分析和验证组样本分类预测

2.3　血清蛋白质/肽指纹图谱分析

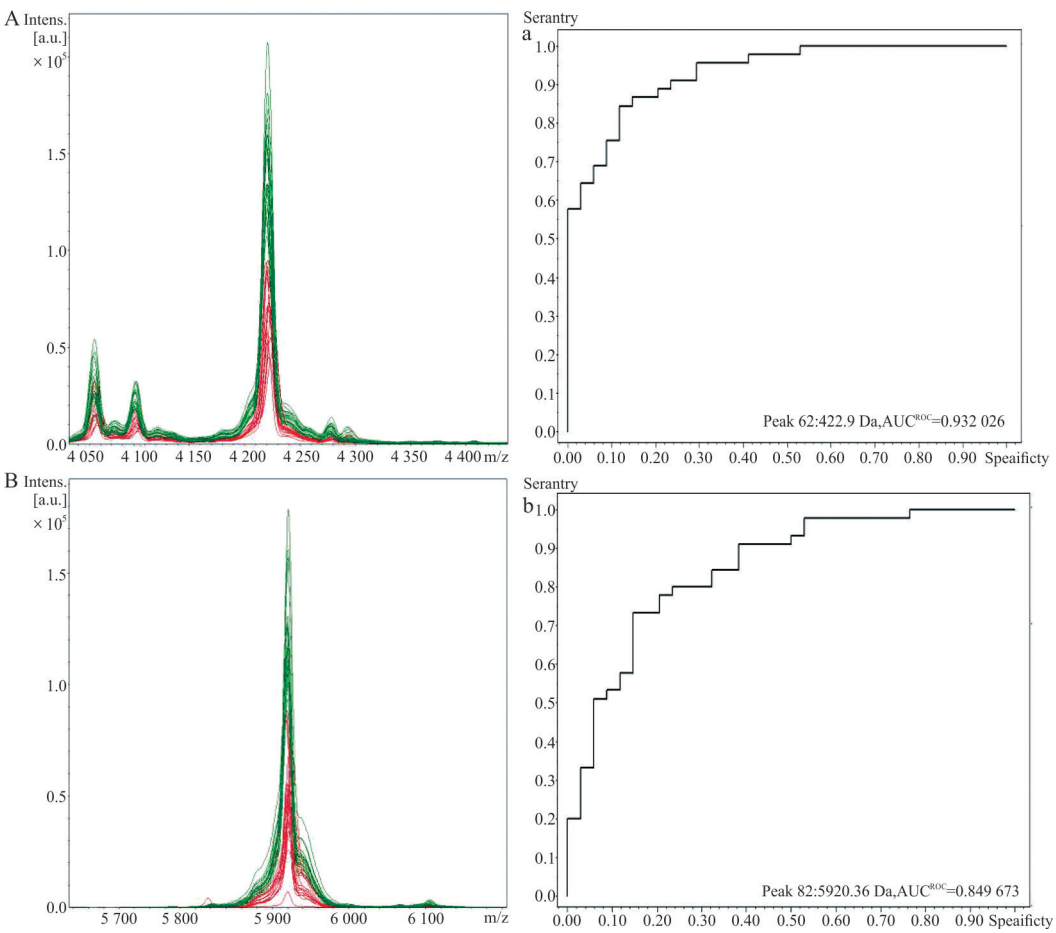
2.3.1　乳腺癌组与对照组图谱分析　通过 ClinProTools 软件分析后，得到乳腺癌组与对照组的血清蛋白质/肽指纹图谱(图 2)，证实两组间图谱存在

明显差异。在分析范围为 1~10 kDa 内发现差异峰 106 个($P<0.05$),筛选出差异有统计学意义($P<0.000\ 001$)且 AUC^{ROC} 值>0.80 的差异峰 26 个。峰 m/z 4 220.9、5 920.36、2 960.40、2 110.95 在乳腺癌组中表达下调,其余 22 个峰均表达上调。此外,m/z 4 220.9、5 920.36 的 2 个峰差异最为显著,其 AUC^{ROC} 值分别为 0.93 和 0.85,且乳腺癌组 m/z 4 220.9、5 920.36 的峰强度明显低于对照组(图 3)。



注:红色为乳腺癌组;绿色为对照组。

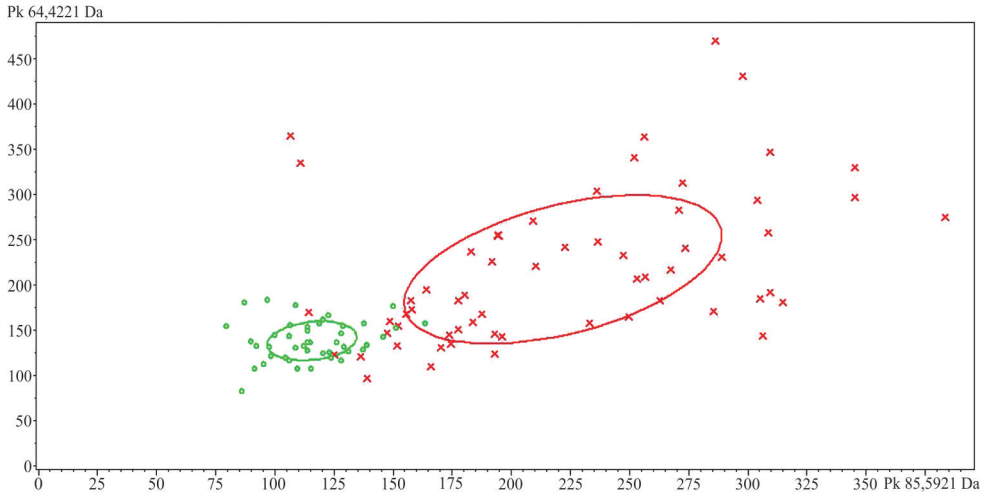
图 2　乳腺癌组与对照组血清蛋白指纹图谱



注:A、a 分别为 m/z 4 220.9 的叠加峰谱图及 ROC 曲线;B、b 分别为 m/z 5 920.36 的叠加峰谱图及 ROC 曲线。红色为乳腺癌组;绿色为对照组。

图 3　差异最显著的 2 个峰叠加峰谱图及其 ROC 曲线

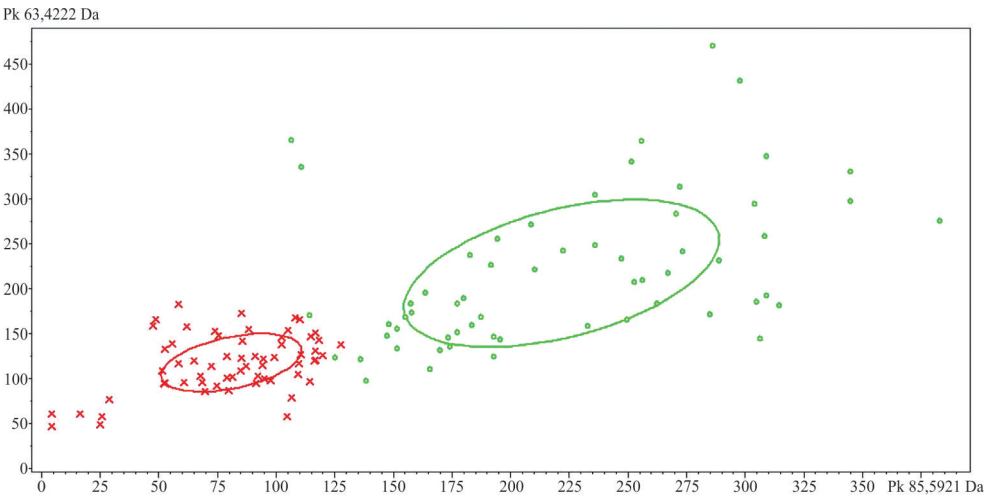
2.3.2　乳腺纤维腺瘤组与对照组图谱分析　经 ClinProTools 软件分析，纤维腺瘤组和对照组两谱图在质量范围 1～10 kDa 之间，存在明显差异，其中具有统计学意义（P<0.05）的差异峰有 40 个。包括 10 个峰表达上调，30 个峰表达下调。此外，m/z 4 220.9、5 920.36 的 2 个峰差异最为显著，其 AUC^{ROC} 值分别为 0.85 和 0.96，且纤维腺瘤组 m/z 4 220.9、5 920.36 的峰强度明显高于对照组。以 m/z 5 920.36、4 220.9 分别为 X、Y 轴进行聚类分析，纤维腺瘤组与对照组可以准确地区分，见图 4。



注：红色叉号为纤维腺瘤组；绿色圆圈为对照组。

图 4　纤维腺瘤组与对照组聚类分析

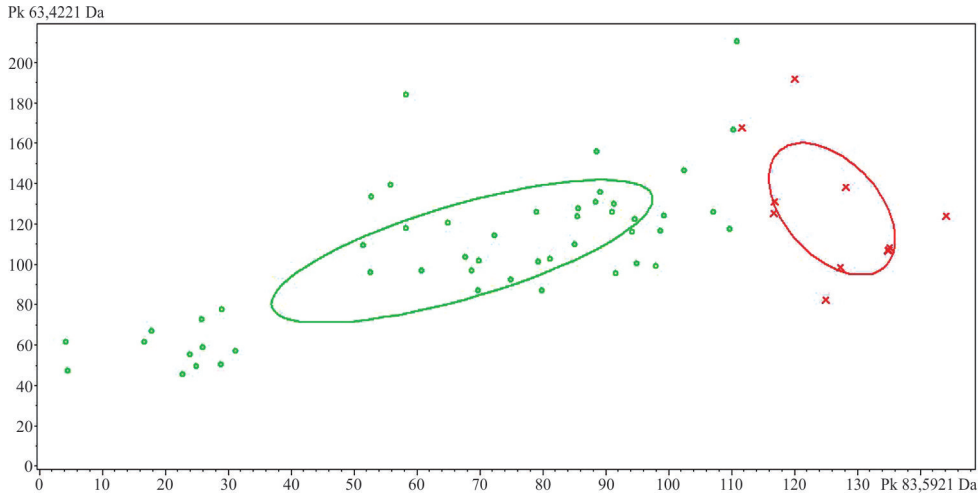
2.3.3　乳腺癌组与纤维腺瘤组图谱分析　经 ClinProTools 软件分析，乳腺癌组与纤维腺瘤组图谱存在明显差异，在 1～10 kDa 范围内发现有统计学意义（P<0.05）的差异峰 78 个，其中 5 个峰表达下调，其余 73 个峰表达上调。此外，m/z 4 220.9、5 920.36 的 2 个峰差异最为显著，其 AUC^{ROC} 值分别为 0.90 和0.99，且乳腺癌组 m/z 4 220.9、5 920.36 的峰强度明显低于纤维腺瘤组。以 m/z 5 920.36、4 220.9 分别为 X、Y 轴进行聚类分析，可准确区分乳腺癌组与纤维腺瘤组（图 5）。



注：红色叉号为乳腺癌组；绿色圆圈为纤维腺瘤组。

图 5　乳腺癌组与纤维腺瘤组聚类分析

2.3.4　TNBC 组与 NTNBC 组图谱分析　经 ClinProTools 软件分析，TNBC 组与 NTNBC 组图谱存在显著差异，在 1～10 kDa 范围内筛选出有统计学意义（P<0.05）的差异峰 40 个，其中 6 个峰表达上调，34 个峰表达下调。此外，m/z 4 220.9、5 920.36 的 2 个峰差异最为显著，其 AUC^{ROC} 值分别为 0.66 和 1，且 TNBC 组 m/z 4 220.9、5 920.36 的峰强度明显高于 NTNBC 组。以 m/z 5 920.36、4 220.9 分别为 X、Y 轴进行聚类分析，可以准确区分 TNBC 和 NTNBC 患者（图 6）。



注：红色叉号为 TNBC 组；绿色圆圈为 NTNBC 组。

图 6　TNBC 组与 NTNBC 组聚类分析

3　讨论

质谱技术作为一种新型蛋白质组学技术，具有高灵敏度、高分辨率、高准确性、高通量等优势，能为疾病的临床诊疗提供快速、精准的分子生物学信息，在医学检验领域受到越来越多的关注。既往大量学者使用 MALDI-TOF-MS 技术对胃癌、肺腺癌、甲状腺乳头状癌、食管鳞癌进行多肽峰的鉴定[3-4]。本研究基于 WCX-MB 纯化技术，应用 MALDI-TOF-MS 对乳腺癌组与对照组血清样本进行分析，采用 3 种算法（GA、SNN、QC）分别建立乳腺癌诊断预测模型并进行验证，结果发现 GA 算法建立的模型诊断效能最佳。GA 算法是一种优化算法，它通过遗传和进化机理，从给出的原始解群中，不断进化产生新的解，直至产生最优解。本研究应用此理论进行取舍，直至得到最佳峰群建立模型。经验证组验证其特异性、敏感性和准确性分别为 90.9%、93.8% 和 92.6%，可有效区分乳腺癌患者与体检健康者。对纤维腺瘤组和对照组、乳腺癌组和纤维腺瘤组以及 TNBC 组和 NTNBC 组分别进行聚类分析结果表明，差异最显著的 2 个质谱峰 m/z 4 220.9、5 920.36 在纤维腺瘤组中的峰强度显著高于对照组；在乳腺癌组中的峰强度显著低于纤维腺瘤组；在 TNBC 组中的峰强度显著高于 NTNBC 组。

综上所述，应用 MALDI-TOF-MS 技术建立的乳腺癌诊断预测模型具有高敏感性、高特异性、高准确性的特点。该技术能够准确区分良恶性乳腺癌和体检健康者，可用于乳腺癌的早期诊断和高危人群筛查，同时可以更直观地区分 TNBC 和 NTNBC，为临床精准治疗提供更为可靠的依据。在乳腺癌分型上，与传统免疫组化染色法相比，MALDI-TOF-MS 技术具有高通量、操作简单以及可排除人为主观因素影响的优点。本次研究筛选出的 2 个差异峰 m/z 4 220.9、5 920.36 有望成为乳腺癌潜在的肿瘤标志物。由于本次实验所收集的样本量较少，研究结果存在偏倚的可能性，因此，后续研究中我们将进一步扩大样本量，纳入更多早期乳腺癌患者，细化分组，丰富实验内容。对于此次获得的差异蛋白质/多肽，后续将使用高效液相色谱-电喷雾串联质谱法进行蛋白质分子鉴定，研究其功能并阐明其在乳腺癌发生发病机制中的作用。

4　参考文献

[1] Taşkın F, Polat Y, Erdoğdu IH, *et al*. Problem-solving breast MRI: useful or a source of new problems? [J]. Diagn Interv Radiol, 2018, 24(5): 255-261.

[2] Lee JW, Lee K, Ahn SH, *et al*. Potential of MALDI-TOF-based serum N-glycan analysis for the diagnosis and surveillance of breast cancer[J]. Sci Rep, 2020, 10(1): 19136.

[3] Lu ZL, Chen YJ, Jing XY, *et al*. Detection and identification of serum peptides biomarker in papillary thyroid cancer[J]. Med Sci Monit, 2018, 24: 1581-1587.

[4] 张婷，王娜娜，杨璐，等. 基于 MALDI-TOF MS 技术食管鳞癌诊断预测模型的建立及初步验证[J]. 临床检验杂志，2019，37(2)：81-86.

（**收稿日期**：2021-07-16）
（**本文编辑**：许晓蒙）