

血清蛋白质组高丰度蛋白去除方法比较研究

赵佳威^{1,2}, 陆 澳^{1,2}, 赵 洋², 孟 波², 叶子弘^{1,2}

(1. 中国计量大学生命科学学院, 浙江省生物计量及检验检疫技术重点实验室, 浙江 杭州 310018;

2. 中国计量科学研究院前沿计量科学中心, 国家市场监督管理总局技术创新中心(质谱), 北京 100029)

摘要: 基于健康人群(HC)和肝癌人群(HCC)的血清样本, 本研究比较高选择 Top14 高丰度蛋白去除试剂盒 (Top14)、DMB 低丰度蛋白富集试剂盒(DMB)2 种血清蛋白质组前处理方案的优劣势。对于 HC 样本, DMB 处理组鉴定的蛋白质数量是 Top14 处理组的 2.6 倍, 是未处理血清原液(Blank)样本的 3.9 倍; 对于 HCC 样本, DMB 处理组鉴定的蛋白质数量是 Top14 处理组的 3.7 倍, Blank 组的 6.2 倍。通过分析蛋白质组的丰度排序散点图, 发现 Top14 预处理对高丰度蛋白质的去除效果显著; 但相较于 DMB 预处理, 低丰度蛋白质的鉴定覆盖率较低。在蛋白质组鉴定频率方面, DMB 处理样本中可定量蛋白质数量明显多于 Top14 处理组和 Blank 组。Top14 处理组和 Blank 组中鉴定的蛋白质超过 50%可在 DMB 处理组中被鉴定到, 且其鉴定蛋白质的 KEGG 分析结果基本可在 DMB 方案中被覆盖。在 DMB 预处理的肝癌血清中鉴定到 47 个差异蛋白质, 而在 Top14 处理组和 Blank 组中分别只鉴定到 15、17 个差异蛋白质。通过对各组差异蛋白质编码基因的 KEGG 分析, DMB、Top14 预处理组和 Blank 样本的差异蛋白质分别富集到 21、5、1 条通路, 表明 DMB 预处理更适用于血清样本蛋白质组的分析。基于 DMB 预处理方法, 本研究发现 HSP90AB1、SPP1、ACTR3、SNCA、PECAM1 和 SRC 在肝癌血清样本中显著上调, 可作为肝癌筛查候选生物标志物。

关键词: 血清; 蛋白质组; 液相色谱-串联质谱(LC-MS/MS)

中图分类号: O657.63

文献标志码: A

文章编号: 1004-2997(2024)04-0519-12

doi: 10.7538/zpxb.2023.0121

Comparative Study on the Methods for Removal of High-Abundant Proteins in Serum Proteome

ZHAO Jia-wei^{1,2}, LU Ao^{1,2}, ZHAO Yang², MENG Bo², YE Zi-hong^{1,2}

(1. Zhejiang Provincial Key Laboratory of Biometrology and Inspection & Quarantine, College of Life Sciences, China Jiliang University, Hangzhou 310018, China; 2. Technology Innovation Center of Mass Spectrometry for State Market Regulation, Center for Advanced Measurement Science, National Institute of Metrology, Beijing 100029, China)

Abstract: This study evaluated the protein identification capabilities of the Top14 high-abundance protein removal kit (Top14) and the DMB MagicOmics low-abundance enrichment kit (DMB) on serum samples from healthy individuals (HC) and hepatocellular carcinoma (HCC) patients. For HC samples, DMB treatment led to identifying 2.6 times proteins more than Top14 and 3.9 times more than untreated samples. For HCC samples, DMB achieved 3.7 times increase in protein identification over Top14 and 6.2 times increase over untreated samples. Although the Top14 kit was effective in

removing high-abundance proteins, it was less proficient in detecting low-abundance proteins when compared to the DMB method. In terms of protein identification frequency, the DMB-treated samples had a significantly higher number of quantifiable proteins than both the Top14 and Blank groups. Over 50% of the proteins identified in the Top14 and Blank groups were also identified in the DMB group, ensuring a comprehensive proteome coverage as evidenced by KEGG analysis. The DMB method significantly outperformed the others in HCC serum samples, identifying 47 differentially expressed (DE) proteins, in contrast to 15 and 17 identified of the Top14 group and untreated samples, respectively, highlighting its superior ability to uncover critical biomarkers for disease analysis. KEGG pathway analysis showed that DE proteins identified by DMB were involved in 21 distinct pathways, significantly more than 5 and 1 pathways identified by Top14 group and untreated samples, demonstrating DMB's advanced proteomic profiling capability. This study also underscored HSP90AB1, SPP1, ACTR3, SNCA, PECAM1, and SRC proteins increased in HCC serum samples based on DMB method, marking them as promising HCC biomarkers for disease screening.

Key words: serum; proteomics; liquid chromatography-tandem mass spectrometry (LC-MS/MS)

血液中蛋白质的变化可以反映机体的病理、生理状态,大部分与疾病相关的生物标志物都与血液中的蛋白质相关^[1]。随着质谱技术的发展,基于质谱的血液蛋白质组学研究已成为疾病早期诊断、新药筛选、生物标志物发现及验证等领域的热点^[2]。2002年,人类蛋白质组组织(Human Proteome Organization, HUPO)正式启动了人类血浆蛋白质组计划(Plasma Proteome Project, PPP)^[3]。血液中的蛋白质成分具有高度的复杂性和多样性,其动态浓度分布范围可高达 10^{12} 数量级^[4]。血液中前10位高丰度蛋白质占据了血浆蛋白总含量的90%,但许多潜在的生物标志物是低丰度蛋白质^[5-6]。在使用质谱等技术进行蛋白质分析时,高丰度蛋白的信号会掩盖低丰度蛋白的信号,严重影响低丰度蛋白的定性和定量分析^[7]。

目前,主要有3种常用的血液样本高丰度蛋白去除方案:1)ProteoMiner蛋白质浓缩试剂盒通过组合肽配体库实现低丰度蛋白浓缩和高丰度蛋白质稀释^[8-10];2)Top14高丰度蛋白去除试剂盒使用高度特异性的固定化抗人血清白蛋白(HSA)和抗免疫球蛋白抗体去除血清和血浆中的HSA以及所有主要的亚类免疫球蛋白^[11-12];3)基于纳米材料富集低丰度蛋白质的方案。2021年,Zhang等^[13]使用ProteoMiner蛋白质浓缩试剂盒处理血浆样本,建立了人血浆样本N-糖蛋白组深度综合分析方案(Glyco-CPLL),集成了组合肽配体库、高pH反相预分级、亲水相互作用色

谱等步骤,从健康人群和甲状腺乳头状癌患者血浆中鉴定了包含786个蛋白质和来自369个N-糖蛋白的1644个独特的完整N-糖肽。2023年,Lu等^[14]使用Top14高丰度蛋白去除试剂盒结合TMT16-plex肽段标记试剂盒,从健康人群、结节性甲状腺肿、甲状腺腺瘤等血浆样本中鉴定了1063个蛋白质。近些年,纳米材料(NPs)被用于富集血液样本中低丰度蛋白质的研究越来越多。2020年,Blume等^[15]将NPs的蛋白质冠层与液相色谱-串联质谱(LC-MS/MS)技术整合,用于高效的蛋白质组学分析,使用5种NPs结合96孔自动工作流程,从141份血浆样本中检测到2000多种蛋白质。2022年,Liu等^[16]报道了载有4种不同基团的二硫化钼纳米材料富集血浆中低丰度蛋白的方案,结合数据非依赖采集(data independent acquisition, DIA)策略,单次质谱分析可鉴定超3400个蛋白质。上述报道介绍了3种不同的血液样本蛋白质组前处理方案,但各自的优劣势尚未有深入的探讨。

ProteoMiner蛋白质浓缩试剂盒对血液样本的需求量较大,常需要多个样本混合使用,无法观察样本间的差异。因此,本研究拟比较Top14高丰度蛋白去除试剂盒(Top14)、基于纳米材料富集血液样本中低丰度蛋白试剂盒(DMB)和血清原液(Blank)在实际临床血液样本蛋白质组分析中的效果。从样本制备方案的稳定性、蛋白质组鉴定深度、低丰度蛋白鉴定效果以及在实际样本差异蛋白质组分析中的应用效果等方面

评价 Top14 和 DMB 试剂盒的特点, 希望给出血液蛋白质组分析试剂盒的选择建议。

1 实验部分

1.1 主要仪器与试剂

Orbitrap Fusion Lumos Tribrid 三合一质谱仪、EASY-nLC 1200 纳升级液相色谱仪、NanoDrop OneC 微量紫外分光光度计、DynaMag-2 磁力架: 美国 Thermo Fisher 公司产品; DHP-3052 电热恒温培养箱: 上海恒一科学仪器有限公司产品; ME104 电子天平: 瑞士 Mettler Toledo 公司产品; Concentrator plus 真空离心浓缩干燥仪: 德国 Eppendorf 公司产品; Milli-Q 型纯水仪、30 ku 超滤管: 美国 Millipore 公司产品。

尿素 (urea, UA)、二硫苏糖醇 (1,4-dithiothreitol, DTT)、碘乙酰胺 (2-iodoacetamide, IAA)、甲酸 (formic acid, FA)、乙腈 (acetonitrile, ACN) 和碳酸氢铵 (NH_4HCO_3): 均为分析纯, 美国 Sigma 公司产品; High Select Top14 Abundant Protein Depletion Mini Spin Columns (Top14, 货号 A36370): 美国 Thermo Scientific 公司产品; MagicOmics Low-abundance Enrichment Kit (DMB, 货号 QLDMB20230504): 北京青莲百奥生物科技有限公司产品; 测序级胰蛋白酶 (货号 P01001): 北京酶知源生物科技有限公司产品。

1.2 样本信息

3 名健康及 3 名肝癌患者的血清样本, 均采集自中国医学科学院、中国协和医科大学, 样本的选择无偏倚。根据中国医学科学院肿瘤医院伦理委员会 (NCC2020C-209) 的批准, 获得所有供试者的知情同意。

1.3 Top14 高丰度蛋白去除及 FASP 酶切方案

取 10 μL 血清于 Top14 旋转柱中, 多次倒置旋转柱直至树脂完全悬浮于溶液中; 将旋转柱放在摇床上, 室温孵育 10 min; 然后拧开旋转柱底部, 以 3 200 r/min 离心 2 min, 收集洗脱液; 使用过滤器辅助样品制备方案 (filter-aided sample preparation, FASP) 酶切^[17-18]。具体而言, 将蛋白转移至 30 ku 超滤管中, 离心去除废液; 加入 200 μL 8 mol/L 尿素 (100 mmol/L Tris-HCl 溶解), 清洗 1 次, 离心去除废液; 向超滤管中加入 100 μL 20 mmol/L DTT 溶液, 37 $^\circ\text{C}$ 还原 4 h, 离心去除废

液; 再加入 100 μL 40 mmol/L IAA, 避光烷基化 30 min, 离心去除废液; 随后加入 100 μL 20 mmol/L DTT 溶液再次反应 15 min, 离心去除废液; 然后分别用 8 mol/L UA 和 50 mmol/L NH_4HCO_3 清洗 3 次; 更换套管, 加入 180 μL 50 mmol/L NH_4HCO_3 溶液, 并加入 20 μL Trypsin 胰蛋白酶进行酶切反应, 37 $^\circ\text{C}$ 孵育 16 h; 反应结束后, 离心收集酶切后的肽段溶液, 45 $^\circ\text{C}$ 真空离心浓缩干燥仪中烘干回收肽段, 于 -80 $^\circ\text{C}$ 冰箱储存, 待 LC-MS/MS 分析。FASP 实验过程中, 均以 12 100 r/min 离心 15 min。Blank 组样本同样使用 FASP 酶切方案。

1.4 DMB 试剂盒富集低丰度蛋白质及蛋白酶切方案

根据制造商的建议, 取 50 μL 血清样本, 使用 DMB 试剂盒富集低丰度蛋白质和溶液酶切方案酶切蛋白质^[19]。具体而言, 用稀释缓冲液稀释样本, 加入磁珠, 37 $^\circ\text{C}$ 孵育 30 min, 在磁力架上静置 2 min 直至澄清, 弃去上清液, 用稀释缓冲液洗涤磁珠 3 次; 用裂解缓冲液代替稀释缓冲液, 95 $^\circ\text{C}$ 加热 10 min, 冷却至室温后, 将磁珠与结合缓冲液室温孵育 15 min, 用洗涤缓冲液清洗 3 次; 替换缓冲体系后, 在 37 $^\circ\text{C}$ 加入胰蛋白酶 (Trypsin) 消化 4 h 后, 用终止缓冲液终止反应; 真空离心浓缩干燥仪干燥肽段, 回收的肽段于 -80 $^\circ\text{C}$ 冰箱储存, 待质谱分析。实验中的试剂均由试剂盒提供。

1.5 实验条件

1.5.1 色谱条件 用微量紫外分光光度计在 205 nm 处测定肽段浓度; 肽段进样量 1 μg , 流速 600 nL/min。自制 C18 毛细管直喷柱 (100 μm \times 30 cm \times 1.9 μm); 流动相: A 为 0.1% FA, B 为 0.1% FA 和 80% ACN; 梯度洗脱程序: 0~6 min (4%~10%B), 6~56 min (10%~22%B), 56~77 min (22%~35%B), 77~88 min (33%~90%B), 88~90 min (90%B)。

1.5.2 质谱条件 高能碰撞解离 (higher-energy collision dissociation, HCD) 碎裂方式, 正离子模式扫描; 一级质谱参数: 质量扫描范围 m/z 350~1 550, 分辨率 120 000, 自动增益控制 (AGC) 4×10^5 , 离子最大注入时间 50 ms; 二级质谱参数: 分辨率 15 000, AGC 为 5×10^4 , 离子最大注入时间 30 ms, 质谱数据由 Xcalibur 实时采集。

1.6 质谱数据处理

使用 Proteome Discoverer(2.3.0.523)和 MaxQuant(2.0.3)软件对质谱原始数据进行定性和定量分析^[20],用于检索的数据库于2023年6月1日从 uniprot 下载,包含20422个蛋白质。Proteome Discoverer 软件参数设置:蛋白水解酶为胰蛋白酶,最大漏切位点2;母离子质量偏差 8×10^6 ,子离子质量偏差 1.8×10^6 ;固定修饰设定半胱氨酸碘乙酰胺化(carbamidomethyl/+57.021 u),可变修饰选择甲硫氨酸氧化(oxidation/+15.995 u)和N-乙酰化(acetyl/+42.011 u);蛋白和多肽的定性假阳性率错误发现率(FDR)为1%,其他参数为默认。MaxQuant 除定量方法选择非标定量(label free)外,其他参数设置与 Proteome Discoverer 软件一致。

1.7 统计学分析

使用 Perseus 软件^[21]完成数据前处理,3次差异分析方案一致。显著性阈值设置 $P < 0.05$,差异倍数(fold change)经过 log2 处理后取 ± 1.2 。特

征鉴定基因、差异基因使用 clusterProfiler4.0^[22]进行 GO 富集分析、京都基因和基因组百科全书(KEGG pathway)^[23]富集分析,选择 $P < 0.05$ 为显著条目。使用 String 网站进行蛋白互作分析(PPI)^[24],使用 Cytoscape 软件优化 PPI 网络。

2 结果分析

2.1 实验方案设计

基于质谱的血液蛋白质组学采用 Top down 质谱采集方案^[25]。从样品中提取蛋白质,将其消化成肽段,再使用质谱法鉴定。然而,由于血液样本中高丰度蛋白质占比超过90%,若不去除,会影响低丰度蛋白质的鉴定^[26]。本研究比较了 Top14 高丰度蛋白去除试剂盒和 DMB 低丰度蛋白富集试剂盒在健康人群(HC)和肝癌人群(HCC)血清样本中的应用效果,示于图1。其中,Blank 和 Top14 采用 FASP 酶切,DMB 采用溶液酶切,使用 LC-MS/MS 分析获得的肽段,MaxQuant 和 Proteome Discoverer 软件解析质谱文件,Perseus 软件分析数据。

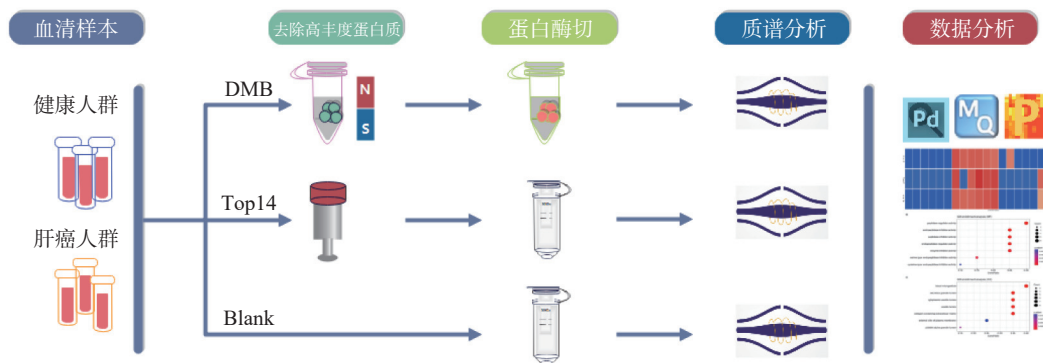


图1 3种血清蛋白质组样本制备方案的技术路线

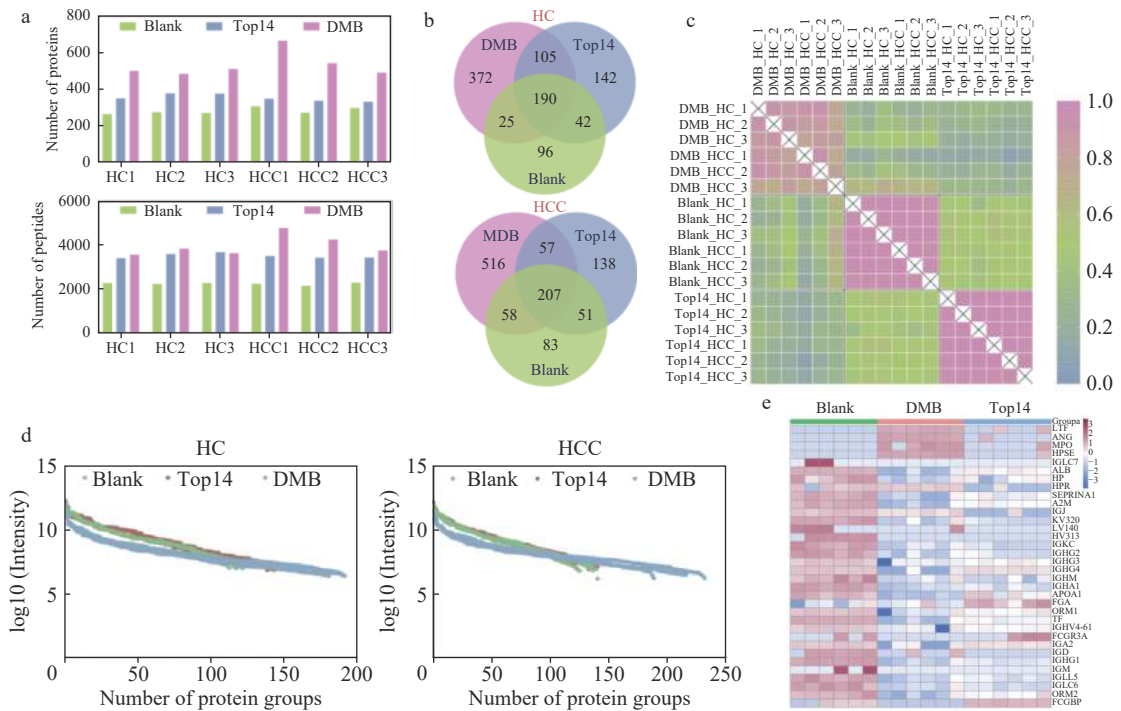
Fig. 1 Technical routes of three sample preparation protocols for serum proteome

2.2 蛋白质和多肽的鉴定

本工作分别采用 DMB、Top14 血清样本前处理试剂盒以及 Blank 对照,完成了3例健康人和3例肝癌患者的血清蛋白质组分析,共鉴定1308个蛋白质和8852条肽段,各样本蛋白质鉴定数在262~660之间,肽段鉴定数在2131~4774之间。在各样本中,Top14处理组的蛋白质和肽段鉴定数量均高于 Blank 组,DMB 处理组鉴定的蛋白质和肽段数量最高,示于图2a。本研究分别在 HC 和 HCC 中对3组鉴定结果进行交

集分析,2类人群3种方案都有特征蛋白质,示于图2b。在 HC 中,DMB 处理组特征蛋白质数量是 Top14 处理组的2.6倍,是 Blank 组的3.9倍;在 HCC 中,DMB 处理组特征蛋白质数量是 Top14 处理组的3.7倍,是 Blank 组的6.2倍。

3组样本的皮尔森相关性分析示于图2c。在 HC 中,DMB 处理组、Blank 组和 Top14 处理组的组内线性相关系数(r)均值分别为0.89、0.96、0.95,3种方案的组内线性相关性良好,但各组间的线性相关性极差。本研究采用数据依



注：a. 各样本制备方案的蛋白质和肽段鉴定数量直方图；b. 3 种样本制备方案在健康人群和肝癌人群中蛋白质鉴定结果的交集分析；c. 所有样本的皮尔森相关性分析；d. 3 种前处理方案鉴定蛋白质组的丰度排序散点图；e. Top14 试剂盒可去除的高丰度蛋白质和血清中部分低丰度蛋白质在 3 种方案中的定量分析热图

图 2 蛋白质和多肽的鉴定

Fig. 2 Identification of proteins and peptides

赖采集 (DDA) 策略, 在一级扫描中选择信号强度高 (TopN) 的离子依次作为母离子进行二级质谱分析。在此模式下, 高丰度抑制效应会导致低丰度蛋白鉴定不稳定, 甚至无法鉴定^[27]。在 HC 和 HCC 中, DMB 处理组、Blank 组和 Top14 处理组鉴定蛋白质组的丰度排序散点图示于图 2d, Top14 处理组和 Blank 组鉴定蛋白质组的丰度相近, DMB 处理组较低。随后, 本研究选取 Top14 试剂盒可去除的 14 种高丰度蛋白质和血清中的部分低丰度蛋白质在 3 组实验中的定量分析结果, 示于图 2e, LTF、ANG、MPO 和 HPSE 为血液中低丰度蛋白质, 其他为高丰度蛋白质。对于高丰度蛋白质, DMB 处理组的定量结果最低, 且多种高丰度蛋白质在多个样本中未被鉴定到; Top14 处理组略高于 DMB 处理组, 但低于 Blank 组, Top14 试剂盒可以达到去除高丰度蛋白质的效果; 对于低丰度蛋白质, 选取了 4 种与肿瘤相关的低丰度蛋白质, 分别是血管生成素 (angiogenin, ANG)、乳铁蛋白 (lactotransferrin, LTF)、髓过氧化物酶 (myeloperoxidase, MPO) 和乙酰肝素酶 (heparanase, HPSE), 这 4 种蛋白质在 Blank 组未

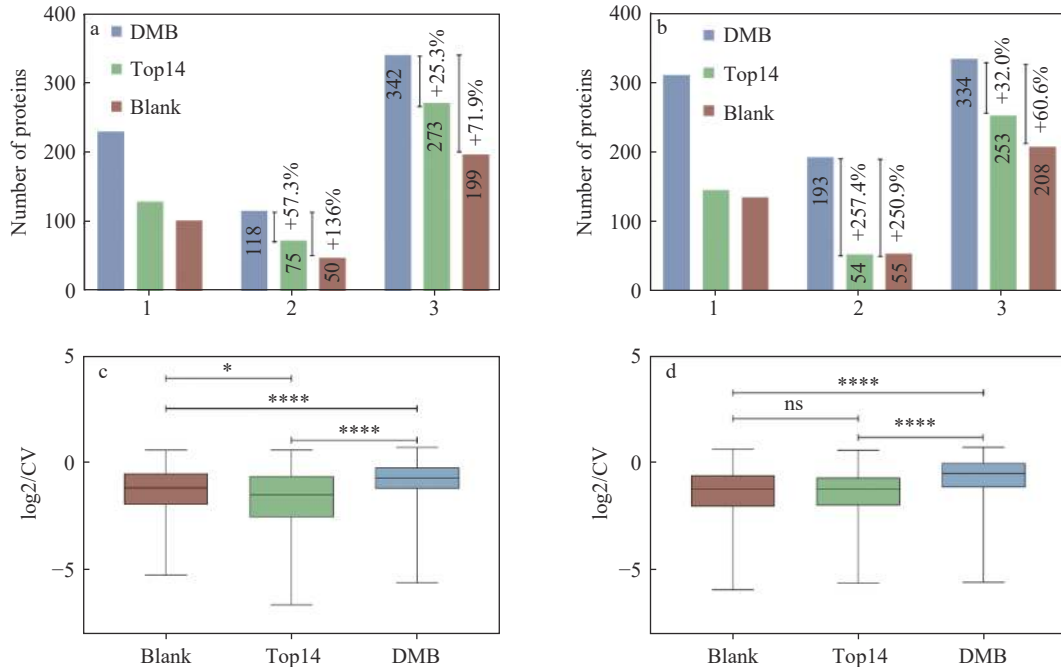
被鉴定到, 在 Top14 处理组仅有几个样本被鉴定到, 但在 DMB 处理组基本均可以被鉴定到。结果表明, DMB 方案在富集血清样本中低丰度蛋白质的效果最佳。

2.3 3 种血清样本前处理方案的稳定性评价

在各组 3 次重复实验中, 成功被鉴定到 2 次及以上的蛋白即可作为定量蛋白, 用于蛋白质组定量分析。统计 Proteome Discoverer 搜库结果中各组鉴定蛋白质的鉴定频率。在 HC 中, DMB 处理组鉴定频率为 2 的蛋白质数量高于其他 2 组, 比 Top14 处理组多 43 个 (57.3%), 比 Blank 组多 68 个 (136%); DMB 处理组鉴定频率为 3 的蛋白质数量同样高于其他 2 组, 比 Top14 处理组多 69 个 (25.3%), 比 Blank 组多 143 个 (71.9%), 示于图 3a。在 HCC 中, DMB 处理组鉴定频率为 2 或 3 的蛋白质数量高于其他 2 组, 增加幅度比 HC 中大, 示于图 3b。这表明, DMB 方案获得可定量蛋白质组的效果更优。从 MaxQuant 搜库结果的 Intensity 矩阵中筛选出可定量蛋白质组, 然后进行中位数归一化和全局最小值插值处理, 再进行变异系数 (coefficient of variation, CV) 分析。

在 HC 和 HCC 中, Blank 组和 Top14 处理组的稳定性更优异, 且差异较小, 示于图 3c、3d。DMB 处理组的稳定性相对较差, 这与鉴定频率为 2 的

蛋白质数量较多有关, 特别是在 HCC 中(图 3b), DMB 处理组鉴定频率为 2 的蛋白质数量是 Top14 处理组的 3.57 倍, 是 Blank 组的 3.51 倍。



注: a. HC 中, 3 种方案各自鉴定蛋白质的频率直方图; b. HCC 中, 3 种方案各自鉴定蛋白质的频率直方图; c. HC 中, 3 种方案鉴定蛋白质组变异系数比较分析; d. HCC 中, 3 种方案鉴定蛋白质组变异系数比较分析; ns 表示 $P>0.05$, *表示 $P<0.05$, ****表示 $P<0.0001$

图 3 3 种血清样本前处理方案的稳定性评价

Fig. 3 Stability assessment of three pre-processing schemes for serum samples

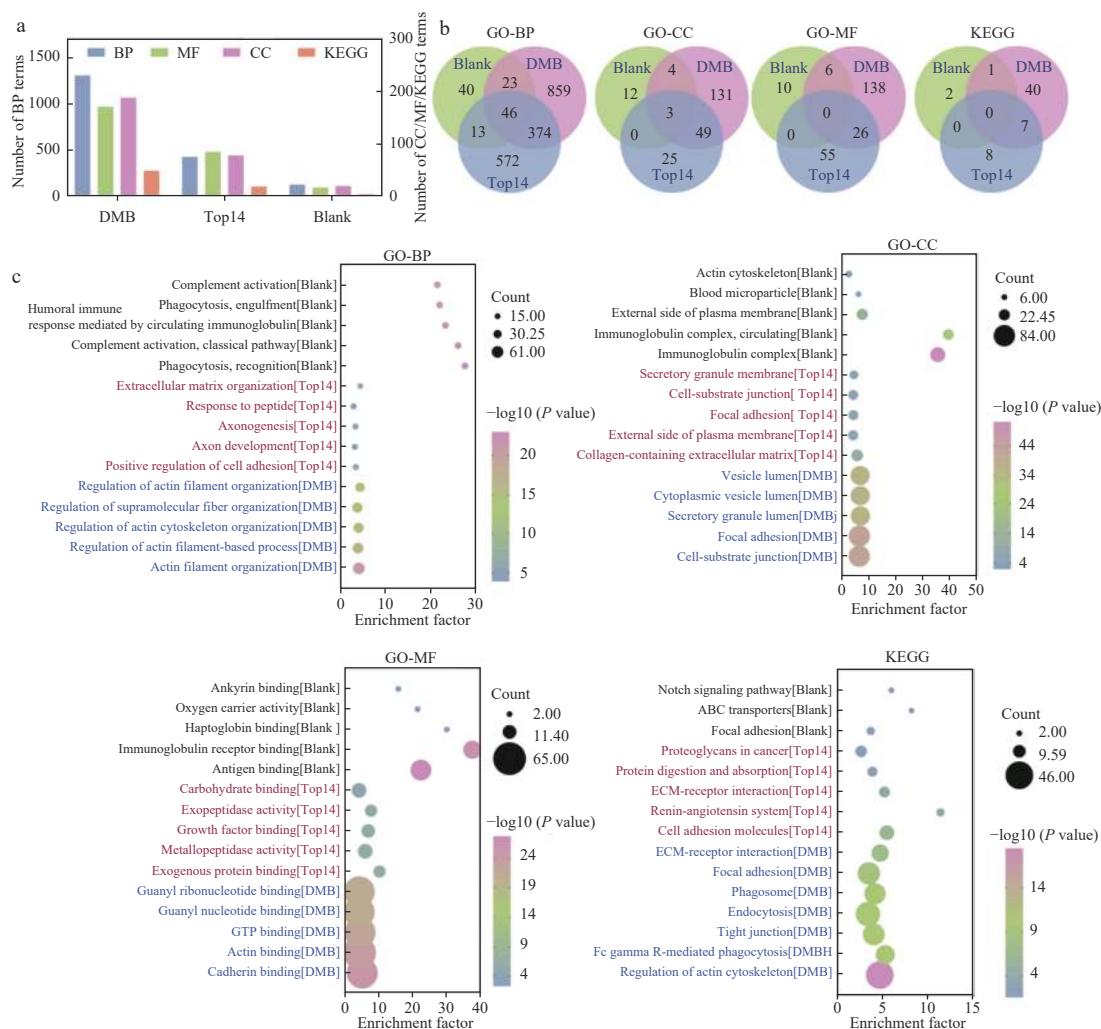
2.4 各方案特征鉴定蛋白质的 GO 和 KEGG 分析

使用 clusterProfile 分别对各组实验在 HC 和 HCC 中鉴定到的特征蛋白质进行 GO 和 KEGG 分析。GO 分析包括生物过程(biological process, BP)、细胞组分(cellular component, CC)和分子功能(molecular function, MF)。富集结果选择 $P<0.05$ 为显著, DMB 处理组特征蛋白质的 GO 和 KEGG 富集结果均最丰富, 相对于 Top14 处理组和 Blank 组有显著变化, 示于图 4a。在各方案特征蛋白质富集结果的交集分析中, 3 组实验均只有少量的交集, DMB 和 Top14 处理组在 GO 和 KEGG 分析结果中均有交集, 但各自也有特征富集的结果, 其中 DMB 处理组特征富集的结果贡献最多, 示于图 4b。这些特点在 KEGG 富集结果中表现得最明显。在参与的生物过程方面, Blank 组特征蛋白质主要参与补体激活和循环免疫球蛋白介导的体液免疫反应等, Top14 处理组特征蛋白质主要参与细胞粘附

的正调控和轴突发育等, DMB 处理组特征蛋白质主要参与基于肌动蛋白丝过程的调控和超分子纤维组织的调控等; 在细胞组分方面, Blank 组特征蛋白质主要来自免疫球蛋白复合体和质膜外侧等, Top14 处理组特征蛋白质主要来自含胶原蛋白的细胞外基质和病灶粘附等, DMB 处理组特征蛋白质主要来自分泌颗粒腔和囊泡腔等; 在分子功能方面, Blank 组特征蛋白质主要有抗原结合和免疫球蛋白受体结合等, Top14 处理组特征蛋白质主要有外源蛋白结合和金属肽酶活性等, DMB 处理组特征蛋白质主要有粘连蛋白结合、肌动蛋白结合和鸟嘌呤-5'-三磷酸(GTP)结合等, 示于图 4c。在 KEGG 通路富集结果方面, Blank 组特征蛋白质主要参与的信号通路有病灶粘附、ABC 转运体和 Notch 信号通路等, 各通路仅有 2~4 个基因; Top14 处理组特征蛋白质主要参与的信号通路有细胞粘附分子、肾素-血管紧张素系统和 ECM-受体相互作用

用等, 各通路中的基因数量虽高于 Blank 组, 但远低于 DMB 处理组; DMB 处理组特征蛋白质

主要参与的信号通路有 ECM-受体相互作用、白细胞跨内皮迁移和肌动蛋白细胞骨架的调节等。



注: a. 富集分析结果条目数直方图; b. 富集分析结果的交集分析; c. 富集分析

图4 各方案特征鉴定蛋白质的GO和KEGG分析

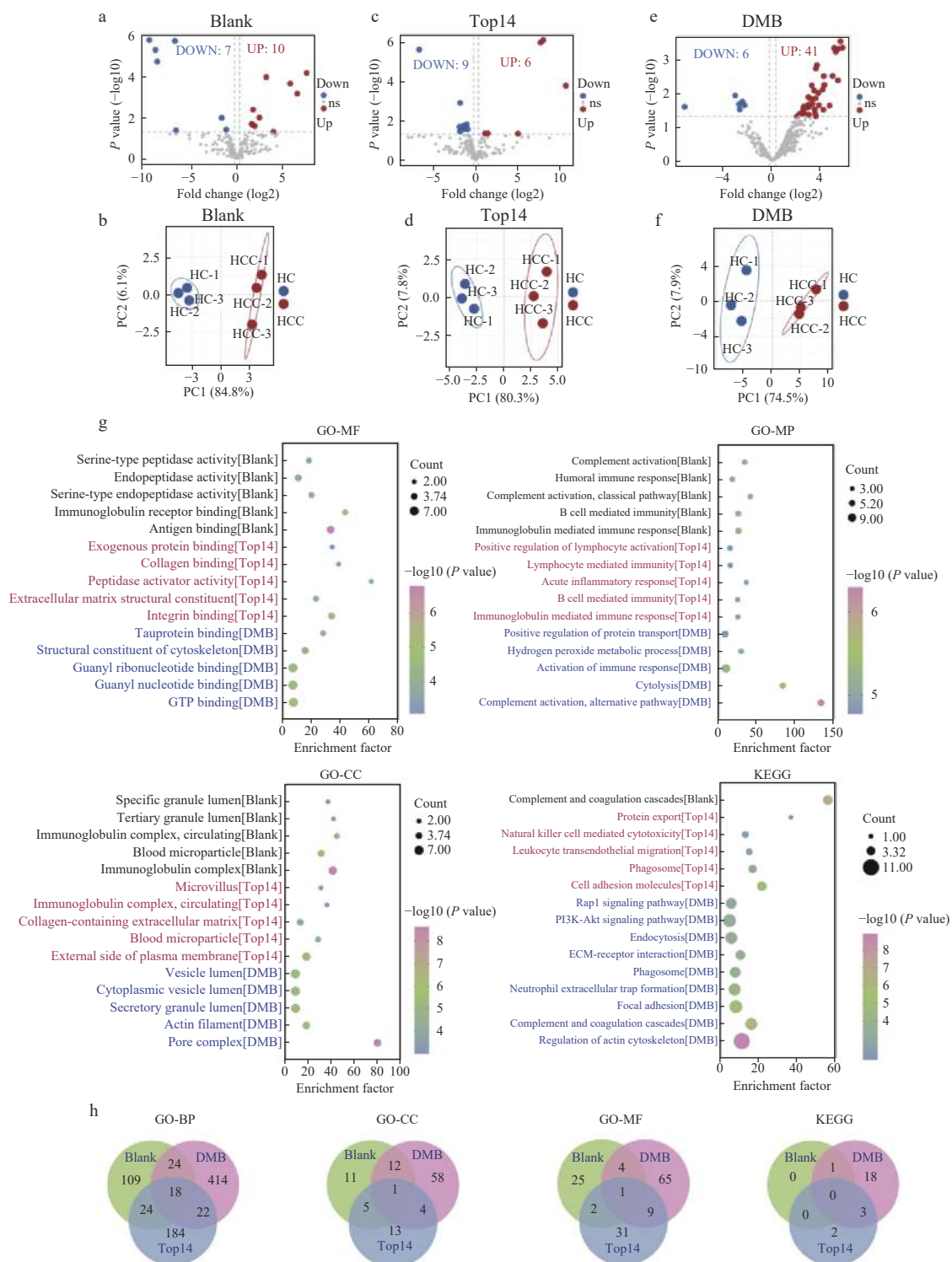
Fig. 4 GO and KEGG analysis of the characteristic identified proteins in each scheme

2.5 健康人群和肝癌人群血清样本差异蛋白质组分析

在我国, 原发性肝癌在恶性肿瘤中排名第四, 在肿瘤致死病因中排名第二, 严重威胁人民的生命和健康^[28]。本研究基于健康人群和肝癌人群的血清样本, 比较了 DMB 和 Top14 血液样本前处理试剂盒相对于血清原液在蛋白质组差异分析中的优劣。

Blank 组共获得 17 个差异基因, 在肝癌人群中 10 个上调, 7 个下调, 示于图 5a。使用 17 个差异基因进行主成分分析 (principal component

analysis, PCA), PC1 为 84.8%, PC2 为 6.1%, 示于图 5b, 可以很好地区分健康人群和肝癌人群。GO 富集结果显示, 差异基因主要来源于免疫球蛋白复合物和血液微颗粒等, 行使的分子功能有抗原结合和免疫球蛋白受体结合等, 参与免疫球蛋白介导的免疫反应和补体激活等生物过程。富集到的 KEGG 通路仅有补体和凝血级联, 示于图 5g。Top14 处理组共获得 15 个差异基因, 在肝癌人群中 6 个上调, 9 个下调, 示于图 5c。使用差异基因进行 PCA 分析, PC1 为 80.3%, PC2 为 7.8%, 示于图 5d, 可以很好地区分健康人群和肝



注: a、c、e 分别为 Blank、Top14 和 DMB 样本制备方案肝癌血清样本差异蛋白质组分析火山图; b、d、f 分别为 Blank、Top14 和 DMB 样本制备方案肝癌血清样本差异表达基因的 PCA 分析; g. 各样本制备方案获得差异表达基因的 GO 和 KEGG 分析; h. 各样本制备方案获得差异表达基因的 GO 和 KEGG 分析结果的交集分析

图 5 Blank、Top14 和 DMB 样本制备方案的肝癌血清样本差异蛋白质组分析
Fig. 5 Analysis of differential protein groups in liver cancer serum samples under Blank, Top14 and DMB sample preparation schemes

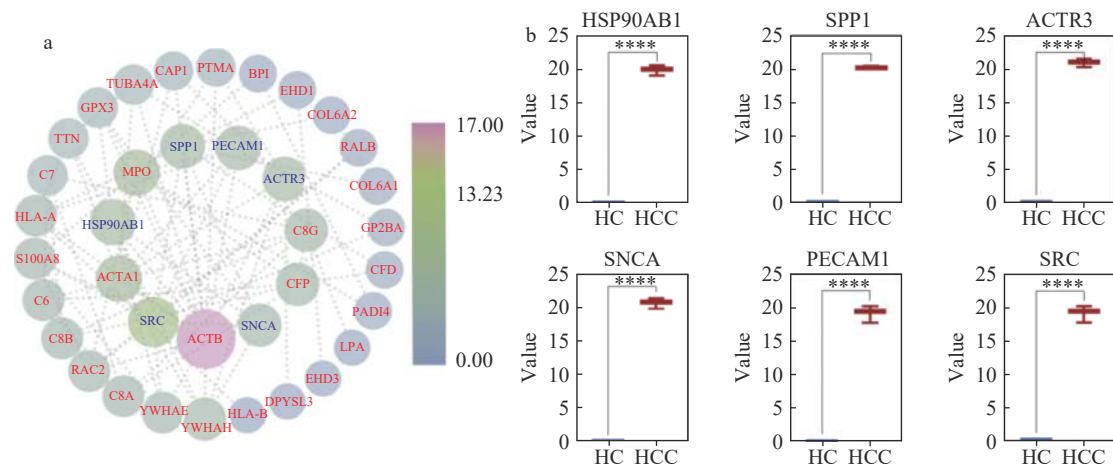
癌人群。GO 富集结果显示,差异基因主要来源于血液微粒和含胶原蛋白的细胞外基质等,行使的分子功能有整合素结合、细胞外基质结构成分和肽酶激活剂活性等,参与免疫球蛋白介导的免疫反应和 B 细胞介导的免疫反应等生物过程。富集的 KEGG 通路有细胞粘附分子和吞噬体等,示于图 5g。DMB 处理组共获得 47 个差异基因,在肝癌人群中表达上调的有 41 个,下调的有 6 个,示于图 5e。差异基因的 PCA 分析可以很好地区分健康人群和肝癌人群,PC1 为 74.5%,PC2 为 7.9%,示于图 5f。富集结果显示,差异基因主要来源于分泌颗粒腔和细胞质囊泡腔等,行使的分子功能有 GTP 结合、肌动蛋白结合等,参与的生物过程有细胞溶解和补体激活等。富集的 KEGG 通路有调节肌动蛋白细胞骨架、补体和凝血级联和病灶粘附等,示于图 5g。

由 3 组实验的差异蛋白质组分析均可发现在肝癌血清样本中显著上调的基因,但获得的差异基因数量和所行使的功能有所不同,这将影响

对肝癌发生发展分子机制的理解。对各组差异基因的 GO 和 KEGG 富集结果进行交集分析后发现,3 组实验之间存在交集,其中 DMB 组中更多的是独自富集到的条目。Blank 仅有 1 条 KEGG 通路,并且可以在 DMB 中富集;Top14 组的 3 条通路也可以在 DMB 中富集,但仅有 2 条通路是特征鉴定的。

2.6 基于 DMB 组差异基因的 PPI 分析

基于上述分析结果,相较于 Blank 组和 Top14 处理组,DMB 处理组对血清中低丰度蛋白的鉴定效果有显著提升,特征鉴定蛋白的功能富集结果更加丰富,发现的差异表达基因数量明显增多。使用 String 对 DMB 组差异基因进行蛋白质互作网络(protein-protein interaction networks, PPI)分析,结合 Cytoscape 对 PPI 网络进行美化,通过 Analyze Network 分析网络,获得 degree 数值,节点的颜色和大小根据 degree 数值变化,示于图 6a。该网络图包含 37 个节点,其中 HSP90AB1、SPP1、ACTR3、SNCA、PECAM1 和 SRC 在健康人群和肝癌人群中表现出显著差异,示于图 6b。



注: a. 差异基因的 PPI 网络图; b. 特征基因的箱线图

图 6 基于 DMB 前处理方案在肝癌血清样本的差异蛋白质组分析

Fig. 6 Differential protein groups analysis in liver cancer serum samples based on the DMB sample preparation scheme

3 讨论

本研究对比了基于 Top14 和 DMB 血液样本前处理试剂盒相对于血清原液在蛋白质组分析中的优劣势。在肽段和蛋白质鉴定方面,DMB 处理组鉴定的蛋白质数量远高于 Top14 处理组和 Blank 组,且对低丰度蛋白的鉴定有更好的结果。相较于 Blank 组,Top14 和 DMB 处理组在

去除高丰度蛋白方面均有效。3 组的组内均有较高相关性,其中 DMB 处理组($r=0.89$)略低于 Top14 处理组($r=0.95$)和 Blank 组($r=0.96$),这与 DMB 处理组鉴定了较多低丰度蛋白质有关。由于高丰度蛋白信号抑制作用,DDA 采集策略会发生采集缺失现象。DMB 特征鉴定蛋白质的 GO 和 KEGG 分析结果远比 Top14 处理组和

Blank 组丰富,证明了 DMB 处理组鉴定蛋白质的多样性。

对健康人群和肝癌人群血清样本的差异蛋白质组进行分析,DMB 处理组获得了 47 个差异基因,是 Top14 处理组的 3.1 倍,Blank 组的 2.8 倍。与 Blank 组和 Top14 处理组的差异基因主要与免疫相关不同,DMB 处理组的差异基因不仅与免疫相关,而且富集到多条与肿瘤相关的信号通路。基于 DMB 处理组获得的差异基因,通过 PPI 分析筛选出多个基因(HSP90AB1、SPP1、ACTR3、SNCA、PECAM1 和 SRC),可用于区分健康人群和肝癌人群。有研究^[29]表明,乳腺癌的生存率和远端转移复发率增加与 HSP90AB1 高表达相关。此外,HSP90AB1 是蛋白质泛素化途径中的伴侣蛋白,对致癌和抑癌基因的调节有重要作用^[30]。2017 年,Xu 等^[31]通过生物信息学分析,发现 SPP1 与上皮-间质转化(EMT)有密切联系,SPP1 在结直肠癌(CRC)组织中的 mRNA 和蛋白表达量明显高于邻近的非癌组织,并且 SPP1 的过表达与 CRC 的肿瘤侵袭、转移和低生存率密切相关。2023 年,Wu 等^[32]通过 qRT-PCR 和免疫组织化学实验验证了 SNCA 在膀胱癌肿瘤组织中下调,并且可能在 T 细胞的浸润方面起重要作用。2023 年,Wu 等^[33]证明了 PECAM-1 缺陷和/或内化是 β -catenin 介导的内皮到间充质细胞转化(EndMT)的关键,而在糖尿病相关的 TME 中,Akt/GSK-3 β 信号通路的增强显著促进了 EndMT 的发生。SRC 是在哺乳动物细胞中第 1 个发现的原癌基因^[34],已发现 SRC 过表达在乳腺癌进展的多种事件中起重要作用,包括肿瘤起始、生长、转移和耐药性等^[35-36]。本研究表明,DMB 血液前处理试剂盒更适用于血清样本蛋白质组分析的样本制备。基于 DMB 试剂盒的肝癌血清生物标志物的筛查,由于样本数量较少,尚需要进一步的验证工作。

参考文献:

- [1] 巩燕,李晓海,应万涛,蔡耘,钱小红. 血液蛋白质组学的研究技术及进展[J]. *生物技术通讯*, 2006, 17(4): 665-668.
GONG Yan, LI Xiaohai, YING Wantao, CAI Yun, QIAN Xiaohong. Techniques for plasma proteome research: progress and prospect[J]. *Letters in Biotechnology*, 2006, 17(4): 665-668(in Chinese).
- [2] 付莉霞,程子倩,王洪,牛明明. 基于质谱的血液蛋白质组学: 血液学研究的新焦点[J]. *中国细胞生物学学报*, 2022, 44(1): 204-213.
FU Lixia, CHENG Ziqian, WANG Hong, NIU Mingming. MS-based blood proteomics: emerging research focus in hematology[J]. *Chinese Journal of Cell Biology*, 2022, 44(1): 204-213(in Chinese).
- [3] OMENN G S, STATES D J, ADAMSKI M, BLACKWELL T W, MENON R, HERMJAKOB H, APWEILER R, HAAB B B, SIMPSON R J, EDDER J S, KAPP E A, MORITZ R L, CHAN D W, RAI A J, ADMON A, AEBERSOLD R, ENG J, HANCOCK W S, HEFTA S A, MEYER H, PAIK Y K, YOO J S, PING P, POUNDS J, ADKINS J, QIAN X, WANG R, WASINGER V, WU C Y, ZHAO X, ZENG R, ARCHAKOV A, TSUGITA A, BEER I, PANDEY A, PISANO M, ANDREWS P, TAMMEN H, SPEICHER D W, HANASH S M. Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database[J]. *Proteomics*, 2005, 5(13): 3 226-3 245.
- [4] SCHIESS R, WOLLSCHIED B, AEBERSOLD R. Targeted proteomic strategy for clinical biomarker discovery[J]. *Molecular Oncology*, 2009, 3(1): 33-44.
- [5] ANDERSON N L, ANDERSON N G. The human plasma proteome: history, character, and diagnostic prospects[J]. *Molecular & Cellular Proteomics*, 2002, 1(11): 845-867.
- [6] TIRUMALAI R S, CHAN K C, PRIETO D A, ISSAQ H J, CONRADS T P, VEENSTRA T D. Characterization of the low molecular weight human serum proteome[J]. *Molecular & Cellular Proteomics*, 2003, 2(10): 1 096-1 103.
- [7] ZHAO Y, XUE Q, WANG M, MENG B, JIANG Y, ZHAI R, ZHANG Y, DAI X, FANG X. Evolution of mass spectrometry instruments and techniques for blood proteomics[J]. *Journal of Proteome Research*, 2023, 22(4): 1 009-1 023.
- [8] BOSCHETTI E, RIGHETTI P G. The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species[J]. *Journal of Proteomics*, 2008, 71(3): 255-264.
- [9] HARTWIG S, CZIBERE A, KOTZKA J, PASSLACK W, HAAS R, ECKEL J, LEHR S. Combinatorial hexapeptide ligand libraries (ProteoMiner): an innova-

- tive fractionation tool for differential quantitative clinical proteomics[J]. *Archives of Physiology and Biochemistry*, 2009, 115(3): 155-160.
- [10] LI L. Dynamic range compression with ProteoMiner™: principles and examples[J]. *Methods in Molecular Biology*, 2015, 1 295: 99-107.
- [11] CAO X, SANDBERG A, ARAÚJO J E, CVETKOVSKI F, BERGLUND E, ERIKSSON L E, PERNEMALM M. Evaluation of spin columns for human plasma depletion to facilitate MS-based proteomics analysis of plasma[J]. *Journal of Proteome Research*, 2021, 20(9): 4 610-4 620.
- [12] KING C D, KAPP K L, ARUL A B, CHOI M J, ROBINSON R A S. Advancements in automation for plasma proteomics sample preparation[J]. *Molecular Omics*, 2022, 18(9): 828-839.
- [13] ZHANG Y, MAO Y, ZHAO W, SU T, ZHONG Y, FU L, ZHU J, CHENG J, YANG H. Glyco-CPLL: an integrated method for in-depth and comprehensive *N*-glyco-proteome profiling of human plasma[J]. *Journal of Proteome Research*, 2020, 19(2): 655-666.
- [14] LU H, PAN Y, RUAN Y, ZHU C, HASSAN H M, GAO J, GAO J, FAN L, LIANG X, WANG H, YING S, CHEN Q. Biomarker discovery for early diagnosis of papillary thyroid carcinoma using high-throughput enhanced quantitative plasma proteomics[J]. *Journal of Proteome Research*, 2023, 22(10): 3 200-3 212.
- [15] BLUME J E, MANNING W C, TROIANO G, HORNBERG D, FIGA M, HESTERBERG L, PLATT T L, ZHAO X, CUARESMA R A, EVERLEY P A, KO M, LIOU H, MAHONEY M, FERDOSI S, ELGIERARI E M, STOLARCZYK C, TANGEYSH B, XIA H, BENZ R, SIDDIQUI A, CARR S A, MA P, LANGER R, FARIAS V, FAROKHZAD O C. Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona[J]. *Nature Communications*, 2020, 11: 3 662.
- [16] LIU Y, YANG Q, DU Z, LIU J, ZHANG Y, ZHANG W, QIN W. Synthesis of surface-functionalized molybdenum disulfide nanomaterials for efficient adsorption and deep profiling of the human plasma proteome by data-independent acquisition[J]. *Analytical Chemistry*, 2022, 94(43): 14 956-14 964.
- [17] ZHAO Y, WANG M, MENG B, GAO Y, XUE Z, HE M, JIANG Y, DAI X, YAN D, FANG X. Identification of dysregulated complement activation pathways driven by *N*-glycosylation alterations in T2D patients[J]. *Frontiers in Chemistry*, 2021, 9: 677 621.
- [18] WIŚNIEWSKI J R, ZOUGMAN A, NAGARAJ N, MANN M. Universal sample preparation method for proteome analysis[J]. *Nature Methods*, 2009, 6: 359-362.
- [19] WANG J, XIE W, SUN L, LI J, WU S, LI R, ZHAO Y. Establishment and clinical application evaluations of a deep mining strategy of plasma proteomics based on nanomaterial protein coronas[J]. *Analytica Chimica Acta*, 2023, 1 275: 341 569.
- [20] COX J, MANN M. MaxQuant enables high peptide identification rates, individualized p.p.b. -range mass accuracies and proteome-wide protein quantification[J]. *Nature Biotechnology*, 2008, 26: 1 367-1 372.
- [21] TYANOVA S, TEMU T, SINITSYN P, CARLSON A, HEIN M Y, GEIGER T, MANN M, COX J. The Perseus computational platform for comprehensive analysis of (pro) omics data[J]. *Nature Methods*, 2016, 13: 731-740.
- [22] WU T, HU E, XU S, CHEN M, GUO P, DAI Z, FENG T, ZHOU L, TANG W, ZHAN L, FU X, LIU S, BO X, YU G. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data[J]. *The Innovation*, 2021, 2(3): 100 141.
- [23] KANEHISA M, GOTO S. KEGG: Kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Research*, 2000, 28(1): 27-30.
- [24] SZKLARCZYK D, FRANCESCHINI A, WYDER S, FORSLUND K, HELLER D, HUERTA-CEPAS J, SIMONOVIC M, ROTH A, SANTOS A, TSAFOU K P, KUHN M, BORK P, JENSEN L J, von MERING C. STRING v10: protein-protein interaction networks, integrated over the tree of life[J]. *Nucleic Acids Research*, 2015, 43(D1): D447-D452.
- [25] CUPP-SUTTON K A, WU S. High-throughput quantitative top-down proteomics[J]. *Molecular Omics*, 2020, 16(2): 91-99.
- [26] ZOLOTARJOVA N, MARTOSELLA J, NICOL G, BAILEY J, BOYES B E, BARRETT W C. Differences among techniques for high-abundant protein depletion[J]. *Proteomics*, 2005, 5(13): 3 304-3 313.
- [27] MEHTA D, SCANDOLA S, UHRIG R G. BoxCar and library-free data-independent acquisition substantially improve the depth, range, and completeness of label-free quantitative proteomics[J]. *Analytical Chemistry*, 2022, 94(2): 793-802.
- [28] 郑荣寿, 张思维, 孙可欣, 陈茹, 王少明, 李荔, 曾红梅, 魏文强, 赫捷. 2016年中国恶性肿瘤流行情况分析[J]. *中华肿瘤杂志*, 2023, 45(3): 212-220.

- ZHENG Rongshou, ZHANG Siwei, SUN Kexin, CHEN Ru, WANG Shaoming, LI Li, ZENG Hongmei, WEI Wenqiang, HE Jie. Cancer statistics in China, 2016[J]. *Chinese Journal of Oncology*, 2023, 45(3): 212-220(in Chinese).
- [29] SUN X, LI K, ARYAL U K, LI B Y, YOKOTA H. PI3K-activated MSC proteomes inhibit mammary tumors via Hsp90ab1 and Myh9[J]. *Molecular Therapy-Oncolytics*, 2022, 26: 360-371.
- [30] HAASE M, FITZE G. HSP90AB1: helping the good and the bad[J]. *Gene*, 2016, 575(2): 171-186.
- [31] XU C, SUN L, JIANG C, ZHOU H, GU L, LIU Y, XU Q. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway[J]. *Biomedicine & Pharmacotherapy*, 2017, 91: 1 167-1 177.
- [32] WU Z, XIA C, ZHANG C, YANG D, MA K. Prognostic significance of SNCA and its methylation in bladder cancer[J]. *BMC Cancer*, 2022, 22(1): 330.
- [33] WU Q, DU X, CHENG J, QI X, LIU H, LV X, GONG X, SHAO C, WANG M, YUE L, YANG X, LI S, ZHANG Y, LI X, YANG H. PECAM-1 drives β -catenin-mediated EndMT via internalization in colon cancer with diabetes mellitus[J]. *Cell Communication and Signaling*, 2023, 21(1): 203.
- [34] YEATMAN T J. A renaissance for SRC[J]. *Nature Reviews Cancer*, 2004, 4(6): 470-480.
- [35] KWAK S M, SEO J, HWANG J T, SUNG G J, SONG J H, JEONG J H, LEE S H, YOON H G, CHOI H K, CHOI K C. EGFR-c-Src-mediated HDAC3 phosphorylation exacerbates invasion of breast cancer cells[J]. *Cells*, 2019, 8(8): 930.
- [36] FINN R S. Targeting Src in breast cancer[J]. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 2008, 19(8): 1 379-1 386.

(收稿日期: 2023-10-23; 修回日期: 2023-12-27)