

doi: 10.12452/j.fxcxb.26072401

不同TNM分期结直肠癌血浆DIA蛋白质组分析及 候选标志物挖掘

张琪^{1,2*}

(1. 北京大学 生命科学学院, 北京 100871; 2. 国家蛋白质科学研究(北京)北京大学分中心, 北京 100871)

摘要: 本研究旨在筛选结直肠癌不同TNM分期血浆差异蛋白, 挖掘无创潜在候选蛋白分子, 为肿瘤早期诊断、分期鉴别、预后评估提供参考, 同时解析肿瘤恶性进展过程的血浆分子变化特征。研究设立健康对照及不同TNM分期结直肠癌血浆队列, 去除血浆高丰度蛋白后, 采用DIA非标记定量蛋白质组技术开展质谱检测, 结合统计分析、功能富集分析, 并依托小规模DIA验证组及TCGA公共数据集开展多维度辅助分析完成数据挖掘。研究共鉴定1736个蛋白群组, 经多组比较交集筛选得到32个上调、19个下调蛋白, 在本研究各TNM分期对比健康对照中均达到差异筛选阈值。富集分析结果显示, 该组上调蛋白富集于凝血活化、肿瘤侵袭、代谢重编程相关条目; 下调蛋白富集条目与免疫调控、肠道黏膜修复相关生物学过程相关。研究还筛选得到9个IV期优势高表达候选蛋白, 其中TIMP1高表达提示患者预后不佳。综上, 本研究解析了结直肠癌进展过程中的血浆蛋白谱紊乱现象, 得到本队列多分期交集蛋白以及晚期优势高表达候选蛋白, 可为结直肠癌无创液体活检、预后分层研究提供候选分子参考。

关键词: 结直肠癌; 血浆; TNM分期; 蛋白质组学; 候选标志物

中图分类号: O657.63; Q51 **文献标识码:** A **文章编号:** 1004-4957(XXXX)XX-0001-11

Plasma DIA Proteomic Analysis of Colorectal Cancer at Different TNM Stages and Mining of Candidate Biomarkers

ZHANG Qi^{1,2*}

(1. School of Life Sciences, Peking University, Beijing 100871, China; 2. National Center for Protein Sciences at Peking University, Beijing 100871, China)

Abstract: This study aims to screen plasma differential proteins in colorectal cancer (CRC) across different TNM stages, identify non-invasive potential candidate protein molecules to provide references for early tumor diagnosis, stage differentiation and prognostic evaluation, and characterize plasma molecular alterations during malignant tumor progression. A plasma cohort consisting of healthy controls and CRC patients at different TNM stages was enrolled. After depletion of high-abundance plasma proteins, mass-spectrometry detection was performed using DIA-based label-free quantitative proteomics. Data mining was implemented by statistical analysis and functional enrichment analysis, with multi-dimensional auxiliary analyses based on a small-scale DIA validation cohort and the TCGA public dataset. A total of 1 736 protein groups were identified. By intersection analysis of multiple-group comparisons, 32 up-regulated and 19 down-regulated proteins were obtained, which satisfied the differential screening thresholds in comparisons between each TNM stage and healthy controls within this cohort. Enrichment analyses revealed that the up-regulated proteins were enriched in items related to coagulation activation, tumor invasion and metabolic reprogramming; the enriched items of down-regulated proteins were associated with biological processes including immune regulation and intestinal mucosal repair. Nine predominantly highly expressed candidate proteins for stage IV were further screened, and high expression of TIMP1 indicated poor patient prognosis. In conclusion, this study characterizes the dysregulated plasma proteomic profiles during CRC progression. The multi-stage intersection proteins and late-stage predominantly highly expressed candidate proteins derived from this cohort can provide candidate molecular

收稿日期: 2026-07-24; 修回日期: 2026-09-01

* 通讯作者: 张琪, 博士, 工程师, 研究方向: 蛋白质组学, E-mail: rubyqq@pku.edu.cn

网络首发日期: XXXX-XX-XX

references for the research on non-invasive liquid biopsy and prognostic stratification of colorectal cancer.

Key words: colorectal cancer; plasma; TNM stage; proteomics; candidate biomarkers

结直肠癌 (Colorectal cancer, CRC) 为全球高发消化系统恶性肿瘤, GLOBOCAN 数据显示其新发病例居恶性肿瘤第三位、死亡病例第二位, 发展中国家发病率持续攀升, 人口老龄化、高脂低纤维饮食、筛查普及率不足是核心诱因 [1]。疾病遵循“腺瘤-癌变”多阶段演进模式 [2], TNM 分期是临床判断肿瘤浸润、淋巴结及远处转移、制定治疗方案与评估预后的金标准 [3], I、II 期局限病灶根治术后 5 年生存率可达 85% 以上, 伴淋巴结转移的 III 期预后显著下滑, IV 期远处转移患者 5 年生存率不足 15% [4]。当前临床筛查手段存在明显短板: 肠镜为诊断金标准但侵入性强、患者依从性差; 血清 CEA 特异性有限, 肠道炎症、良性息肉均可造成假阳性, 近 30% 早期 CRC 无 CEA 升高, 且无法精准区分不同进展分期 [5]。同一 TNM 分期内部患者预后异质性突出, 部分 II 期病例术后短期内复发, 传统病理指标难以解释个体转移潜能差异, 临床亟需无创、可分层分期的循环分子标志物, 通过血浆蛋白组刻画肿瘤渐进式侵袭特征, 对完善 CRC 精准诊疗体系具备重要临床价值 [6]。

蛋白质是细胞生理病理功能的直接执行载体, 相比基因组、转录组更能实时反映肿瘤动态表型, 质谱定量蛋白质组已成为肿瘤标志物挖掘与发病机制解析的核心工具 [7, 8]。传统定量质谱分为标记定量与非标记 DDA、基于谱图库的 DIA 三类技术: 标记定量试剂成本高、通道受限; DDA 仅随机碎裂高丰度肽段, 低丰度蛋白检出稳定性差, 样本间定量重复性不足; 基于谱图库的 DIA 需提前构建特异性肽段图谱库, 增加实验批次偏差与样本损耗 [9]。DirectDIA 作为新一代无文库 DIA 定量策略, 无需单独制备建库样本, 直接依托一级母离子窗口并行碎裂全部洗脱肽段, 兼顾蛋白鉴定深度、定量重现性与检测通量, 适配临床微量血浆样本检测场景 [10]。该技术可均匀覆盖宽质量数区间肽段, 稳定捕获血浆低丰度肿瘤分泌蛋白、免疫调控蛋白与基质重塑分子 [11], 近年已用于肺癌、肝癌体液蛋白组筛选, 但覆盖 CRC 完整 TNM 分期梯度的系统性 directDIA 血浆蛋白组研究仍较为稀缺, 缺乏从早期局限灶至全身转移的连续定量数据, 难以阐明蛋白谱随肿瘤浸润、转移递进的动态调控规律。

血浆是临床最易获取的无创液体活检载体, 富集肿瘤细胞分泌蛋白、微环境渗漏蛋白、机体循环免疫相关分子, 是发掘肿瘤分期标志物理想材料, 但血浆蛋白浓度动态范围跨越 10 - 12 个数量级, 白蛋白、免疫球蛋白等高丰度蛋白占总蛋白 95% 以上, 大量潜在低丰度肿瘤特征蛋白极易被高丰度信号掩盖, 大幅提升差异分子筛选难度 [12]。现有 CRC 血浆蛋白组研究多聚焦健康人群与早期肿瘤二元对比, 仅挖掘早期诊断标志物, 忽略肿瘤进展过程中的梯度分子变化, 难以区分局部进展与远处转移病例; 多数研究采用 DDA 或低分辨质谱, 筛选得到的候选蛋白在独立队列中分期判别效能不佳 [13]。血浆蛋白质组在 CRC 领域具备三大核心应用方向: 一是无创辅助 TNM 分期, 弥补病理与传统血清标志物分层缺陷; 二是转移风险预判, 筛选介导侵袭、远处定植的循环分泌蛋白, 识别高复发亚群; 三是机制解析, 阐释肿瘤进展中免疫抑制、细胞外基质重塑、血管生成的体液分子通路变化 [14]。已有研究证实 CRC 发生淋巴结、远处转移时, 基质金属蛋白酶、趋化因子、EMT 相关循环蛋白呈梯度表达改变 [15], 但尚未依托 directDIA 技术系统绘制 I-IV 期完整血浆蛋白动态图谱, 缺少可精准匹配 TNM 进展梯度的特征蛋白集合。

综合现有研究局限, 本研究纳入不同 TNM 分期 CRC 患者血浆样本, 采用 directDIA 无文库非标记定量蛋白质组技术开展血浆蛋白全景定量。directDIA 技术在肿瘤蛋白质组领域已有相关报道, 本研究重点在临床研究视角与研究设计层面开展探索。既往多数结直肠癌血浆 DIA 蛋白质组研究多采用健康-肿瘤二元分组, 较少完整覆盖 I-IV 全部 TNM 分期梯度, 缺少对肿瘤由局部浸润向远处转移演进过程中血浆蛋白动态变化的系统刻画。本研究系统对比 I、II、III、IV 期全梯度血浆蛋白表达谱, 筛选随肿瘤进展发生梯度改变的差异蛋白集合, 结合生物信息学解析上述蛋白关联的侵袭、免疫及转移相关通路, 区分得到跨队列多分期存在差异的蛋白集合以及 IV 期优势高表达的转移相关候选分子, 实现广谱诊断候选分子与晚期转移预警候选分子的分类挖掘。本研究依托 directDIA 稳定性好、样本损耗低的技术优势, 完整覆盖结直肠癌从早期局限病灶到远处转移的疾病进展周期; 结合小规模 DIA 验证组以及 TCGA 公共数据库完成多维度辅助分析。受临床样本获取条件限制, 本研究属于标志物挖掘的探

索性工作,旨在从上千个血浆蛋白中压缩得到数量有限的候选分子清单,为后续更大样本独立队列靶向验证以及体内外功能机制研究缩小筛选范围、节约科研成本。同时本研究产出一套中国人群结直肠癌不同 TNM 分期的 directDIA 血浆蛋白组实测数据集,可为国内同领域后续研究提供参考组学资源。

1 实验部分

1.1 主要仪器与装置

台式低温离心机、Vanquish Neo 液相色谱、Orbitrap Astral 超高分辨质谱仪、恒温培养箱:美国 Thermo Fisher Scientific 公司产品;真空离心浓缩仪、恒温混匀仪:德国 Eppendorf 公司产品;酶标仪:美国 BioTek 公司产品;漩涡混匀仪:美国 Scilogex 公司产品;电子天平:德国 Sartorius 公司产品。

1.2 主要材料与试剂

Pierce™ BCA 蛋白定量试剂盒、High-Select™ Top14 高丰度蛋白去除树脂试剂盒、质谱级乙腈 (ACN)、质谱级甲酸 (FA)、质谱级水:美国 Thermo Fisher Scientific 公司产品;尿素 (urea)、二硫苏糖醇 (DTT)、碘乙酰胺 (IAM)、三氟乙酸 (TFA):美国 Sigma-Aldrich 公司产品;测序级胰蛋白酶 (Trypsin)、Lys-C 内切蛋白酶:美国 Promega 公司产品。

1.3 试验条件

1.3.1 血浆采集及标本保存 健康对照人群与 CRC 患者血液样本均取自首都医科大学附属北京朝阳医院。患者纳入标准:经病理确诊为结直肠癌;采血前未接受放化疗、靶向治疗或免疫治疗等抗肿瘤相关治疗。排除标准:合并其他恶性肿瘤、严重肝肾功能不全、自身免疫性疾病、急慢性感染性疾病。健康对照纳入标准:同期于本院体检的健康志愿者;无消化道肿瘤病史。排除标准:患有炎症性肠病、结直肠息肉、其他恶性肿瘤、严重全身性疾病及感染性疾病。研究方案经医院伦理委员会审批 (No. 2021-3-1-17),受试者均签署知情同意书。清晨采集空腹静脉血 5 mL,柠檬酸盐与血液按照体积比 1:9 混合抗凝,10 °C、1800 g 条件下离心 10 min,吸取上层血浆分装,置于-80 °C 冰箱保存,全程避免样本反复冻融。

1.3.2 血浆样本 Top14 高丰度蛋白质的去除 将去高丰度蛋白树脂柱放置至室温平衡,取 10 μL 血浆样本加入树脂柱内,密封柱体后反复颠倒,使树脂与样品充分混匀;室温轻柔颠倒孵育 10 min,充分结合白蛋白、IgG 等高丰度蛋白;孵育结束后打开树脂柱上下端盖子,置于 2 mL 收集管中,1000 g 离心 2 min,弃去树脂柱,管中滤液即为去除高丰度蛋白后的蛋白溶液。

1.3.3 蛋白质还原、烷基化、酶解 BCA 试剂盒测定蛋白溶液浓度,各组取等量蛋白 50 μg,使用 8 M urea 将体系补足至 20 μL;加入 DTT 至终浓度 5 mM,室温孵育 30 min 完成蛋白还原;加入 IAM 至终浓度 10 mM,室温避光孵育 30 min 封闭蛋白巯基;再次添加 DTT 至终浓度 5 mM,室温孵育 10 min 中和过量 IAM;按照蛋白与 Lys-C 质量比 100:1 加入蛋白酶,37 °C 酶解 2 h;添加 25 mM 碳酸氢铵稀释 urea 至浓度低于 2 M;按照蛋白与胰蛋白酶质量比 50:1 加入测序级 Trypsin,封口膜密封后 37 °C 恒温酶解 8~18 h;最后加入终浓度 0.1% TFA,终止酶解反应,根据肽段总量选择适配脱盐柱脱盐,真空离心浓缩仪中完全抽干备用。

1.3.4 色谱条件 发现组采用 0.1% FA 水溶液重溶干燥肽段,18000 g 离心 15 min,吸取上清转移至质谱进样管;液相系统选用 Thermo Vanquish™ Neo UHPLC 超高效液相,色谱柱为 PepMap™ RSLC C₁₈ 色谱柱 (2 μm, 100 Å, 150 μm×15 cm),流动相 A 相为 0.1% FA 水溶液, B 相为 80%ACN/0.1% FA 溶液,总洗脱时长 13 min,洗脱程序如下: 0~0.7 min, B 相比比例 1%~4%, 流速 1.8 μL/min; 0.7~1.0 min, B 相 4%~8%, 流速 1.8 μL/min; 1.0~7.7 min, B 相 8.5%~22.5%, 流速 1.8 μL/min; 7.7~11.4 min, B 相 22.5%~45%, 流速 1.8 μL/min; 11.4~11.8 min, B 相 45%~65%, 流速 2.5 μL/min; 11.8~12.3 min, B 相 65%~99%, 流速 2.5 μL/min; 12.3~13.0 min, 维持 B 相 99%, 流速 2.5 μL/min。验证组多肽样品采用 EASY-nLC 1200 液相色谱系统进行分离,将样品上样至 C₁₈ 色谱体系,预柱选用 Acclaim PepMap™ 100 (75 μm × 2 cm, nanoViper 2Pk, C₁₈, 3 μm, 100Å),分析柱为 Acclaim PepMap™ RSLC (75 μm × 25 cm, nanoViper, C₁₈, 2 μm, 100Å)。流动相 A 为 0.1% FA 水溶液, B 相为 80%ACN/0.1% FA 溶液,流速设置为 300 nL/min,总洗脱梯度时长 60 min;梯度程序如下:

0-2 min, B相比比例4%-8%; 2-39 min, B相8%-25%; 39-50 min, B相25%-35%; 50-57 min, B相35%-95%; 57-60 min维持B相95%进行洗脱。

1.3.5 质谱条件 发现组质谱检测采用Thermo Orbitrap Astral质谱仪, 配套Xcalibur软件控制, 以DIA模式采集数据; EASY-SPRAY离子源, 喷雾电压设置2.3 kV, 离子传输管温度320 °C, MIPS模式设为Peptide。一级MS1扫描参数: Orbitrap分辨率240000 ($m/z=200$), 质荷比扫描范围380~980, RF透镜比例40%, AGC目标值 5×10^6 , 最大离子注入时间5 ms。二级DIA碎裂扫描参数: 母离子扫描区间380~980, 隔离窗口宽度2 m/z , 窗口无重叠, 共设置299个扫描窗口, HCD归一化碰撞能量25%, 检测器为Astral, 碎片离子扫描范围150~2000 m/z , AGC目标值 5×10^4 , 最大离子注入时间3 ms。验证组分离后的肽段经由Nanospray Flex离子源, 在Thermo Orbitrap Exploris 480质谱仪上开展DIA模式数据采集。MS1全扫描Orbitrap分辨率设置为60000 (@ m/z 200), 扫描 m/z 范围350-1200, RF Lens为50%, AGC Target设为 3×10^6 , 最大注入时间50 ms; DIA二级扫描母离子 m/z 范围400-1000, 隔离窗口宽度10 m/z , 窗口重叠为0 m/z , 共设置60个扫描窗口, HCD归一化碰撞能量NCE为32%, 二级谱图Orbitrap分辨率30000 (@ m/z 200), 子离子扫描 m/z 范围200-2000, AGC Target为 2×10^6 , 最大注入时间54 ms。

1.3.6 质谱数据检索 质谱原始raw文件导入Spectronaut 19.9软件, 采用Pulsar搜索引擎执行direct-DIA检索。检索数据库选用Uniprot UP000005640_9606_Homo sapiens参考蛋白库, 酶切方式为胰蛋白酶特异性酶切, 允许最多2个漏切位点; 肽段长度限制为7~52个氨基酸, 仅识别b、y型碎片离子; 半胱氨酸烷基化设置为固定修饰, 甲硫氨酸氧化、蛋白N端乙酰化设为可变修饰; 蛋白、肽段、谱图水平False Discovery Rate (FDR)均控制在1%以内; 检索质荷比区间200~3000 m/z , 每个母离子匹配3~6条最优碎片离子, 基于MS2峰面积完成蛋白定量。

1.3.7 生物信息学与统计学分析 质谱定量数据在RStudio软件中完成生物信息学与统计分析。首先对蛋白质定量数据矩阵进行缺失值的筛选以及KNN近邻法(k-Nearest Neighbor)的缺失值填充, 进一步在蛋白水平对数据进行中位数标准化校正, 数据经log转换后采用单因素方差分析筛选组间差异表达蛋白并对假设检验结果绘制火山图; 针对差异蛋白开展Gene Ontology (GO)功能注释与Kyoto Encyclopedia of Genes and Genomes (KEGG)信号通路富集分析; 通过Principal component analysis (PCA)主成分分析直观展示各组样本蛋白表达谱的区分程度; 采用Kaplan-Meier生存曲线结合log-rank检验, 分析蛋白表达水平与CRC患者生存预后的关联。本研究directDIA质谱检测未设置技术重复, 全部为生物学重复样本。

2 结果与讨论

2.1 结直肠癌血浆临床样本的分组

发现组24例血浆样本包括4例健康者和20例CRC患者的血浆, 并对患者的病理分期进行了详细的报告, 根据肿瘤浸润情况(T)、区域淋巴结受累(N)和远处器官转移(M)三个肿瘤特征要素来制定的TNM分期原则, 将患者划分为TNM I期5例, TNM II期5例, TNM III期5例和TNM IV期5例。验证组15例血浆样本包括健康者和4个不同TNM时期结直肠癌患者各3例。

2.2 血浆蛋白质组统计分析

血浆蛋白质组研究受到巨大的跨10个数量级的蛋白质丰度动态范围的限制, 许多感兴趣的蛋白质在血浆中含量较低, 在高丰度的蛋白质中很难检测到。本研究采用Top14去高丰度蛋白试剂盒去除约占人血浆样品中蛋白质95%的高丰度蛋白, 并采用质谱鉴定和定量样品中的中、低丰度蛋白。对所有去高丰度蛋白后的血浆样本进行BCA蛋白浓度测定并取等质量的蛋白质进行溶液内酶解, 酶解后进行肽段脱盐。

对24例样品通过不依赖于谱图库的directDIA检索方式的数据检索, 发现在所有血浆样品中一共可以鉴定到1736个蛋白群组(图1A), 定量蛋白质及相关数值列于附表1。比较24例样品中蛋白质群组鉴定的情况, 发现在所有样品中都鉴定到的蛋白群组为904个, 平均每个样品鉴定到1505个蛋白群组, 说明来自不同人的血浆样品蛋白的异质性很高。观察鉴定定量到的血浆蛋白质的丰度分布情况, 发现

去高丰度蛋白后的血浆样品蛋白质的丰度动态范围跨越6个数量级, 但大多数蛋白的表达量在4个数量级的范围里(图1B)。本研究质谱鉴定结果覆盖多种公认CRC相关功能蛋白, 包含临床常规检测标志物CEA, 该分子广泛用于CRC筛查、术后复发监测与预后评估^[16]; 同时检出MMP9、EGFR、CA72-4、CA242等具备明确临床功能的蛋白: MMP9可重塑细胞外基质、促进肿瘤侵袭转移, 是CRC独立不良预后标志物, 同时也是潜在转移干预靶点^[17]; EGFR是RAS野生型转移性CRC的核心治疗靶点, 其表达水平可预判抗EGFR单克隆抗体药物的治疗响应效果^[18]; CA72-4和CA242为辅助型血清糖蛋白标志物, 与CEA联合检测可大幅提升早期CRC诊断准确率, 同时能够预判肿瘤转移风险^[19, 20]。多种经临床验证的特征蛋白在本数据集中稳定检出, 证实本次蛋白质组质谱数据可靠、分子鉴定真实有效, 数据集内大量未被临床开发的差异蛋白也为后续新型标志物挖掘、靶向分子机制研究提供充足数据基础。

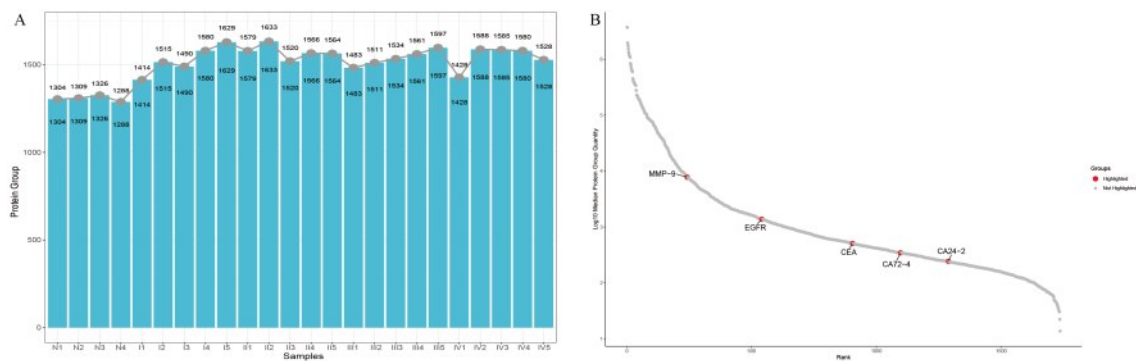


图1 血浆蛋白质群组鉴定数量(A)和丰度动态范围(B)

Fig. 1 Number of identified plasma protein groups (A) and dynamic range of abundance (B)

基于所鉴定到的全部血浆蛋白质, 对24例样本进行PCA主成分分析, 发现健康人的血浆样本可以与CRC患者的血浆样本明显区分, 不同TNM时期患者的血浆样本无法明显区分(图2A)。进一步对24例样本进行中位数标准化处理, 将每个样本的尺度拉到同一水平, 使得数据更具有可比性(图2B)。

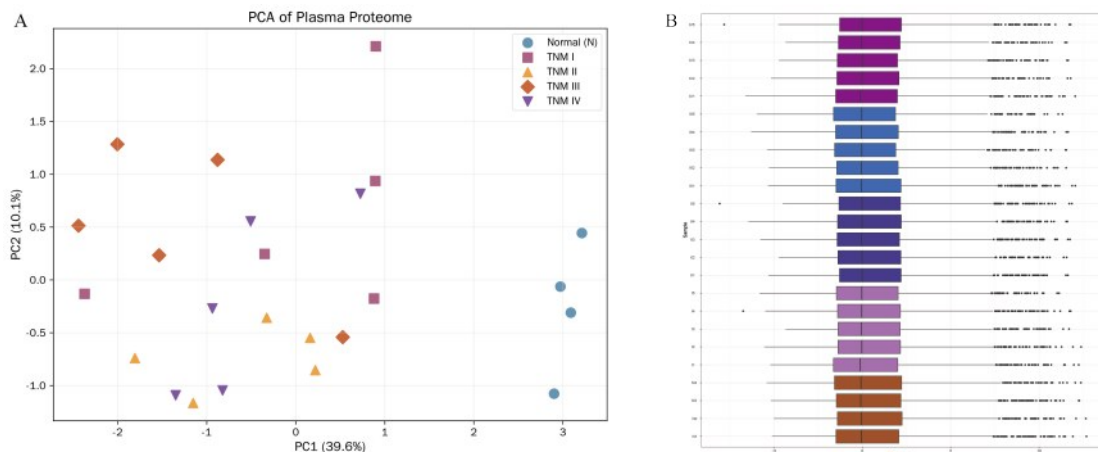


图2 血浆样本主成分分析(A)与数据标准化后分布箱线图(B)

Fig. 2 PCA (A) and boxplots of distribution after data normalization (B) of plasma samples

2.3 结直肠癌血浆中的差异表达蛋白筛选与功能富集分析

为了寻找CRC患者与健康人血浆中差异表达的蛋白, 将不同TNM分期患者分别与健康人进行单因素方差分析的假设检验, 将变化倍数在1.5倍以上, Benjamini-Hochberg (BH)校正的 adjust p -value 小于0.05的蛋白筛选为差异表达蛋白。火山图展示了各个时期差异表达的蛋白情况(图3), 在TNM I期有35个上调蛋白, 45个下调蛋白; II期有144个上调蛋白, 112个下调蛋白; III期有164个上调蛋白, 240个下调蛋白; IV期有151个上调蛋白, 154个下调蛋白, 上/下调表达蛋白序号及其变化倍数详见附表2。相比而言, TNM I期鉴定到的差异蛋白数量更少, 可能是结肠癌患者早期血浆蛋白的变化没有那么剧烈。

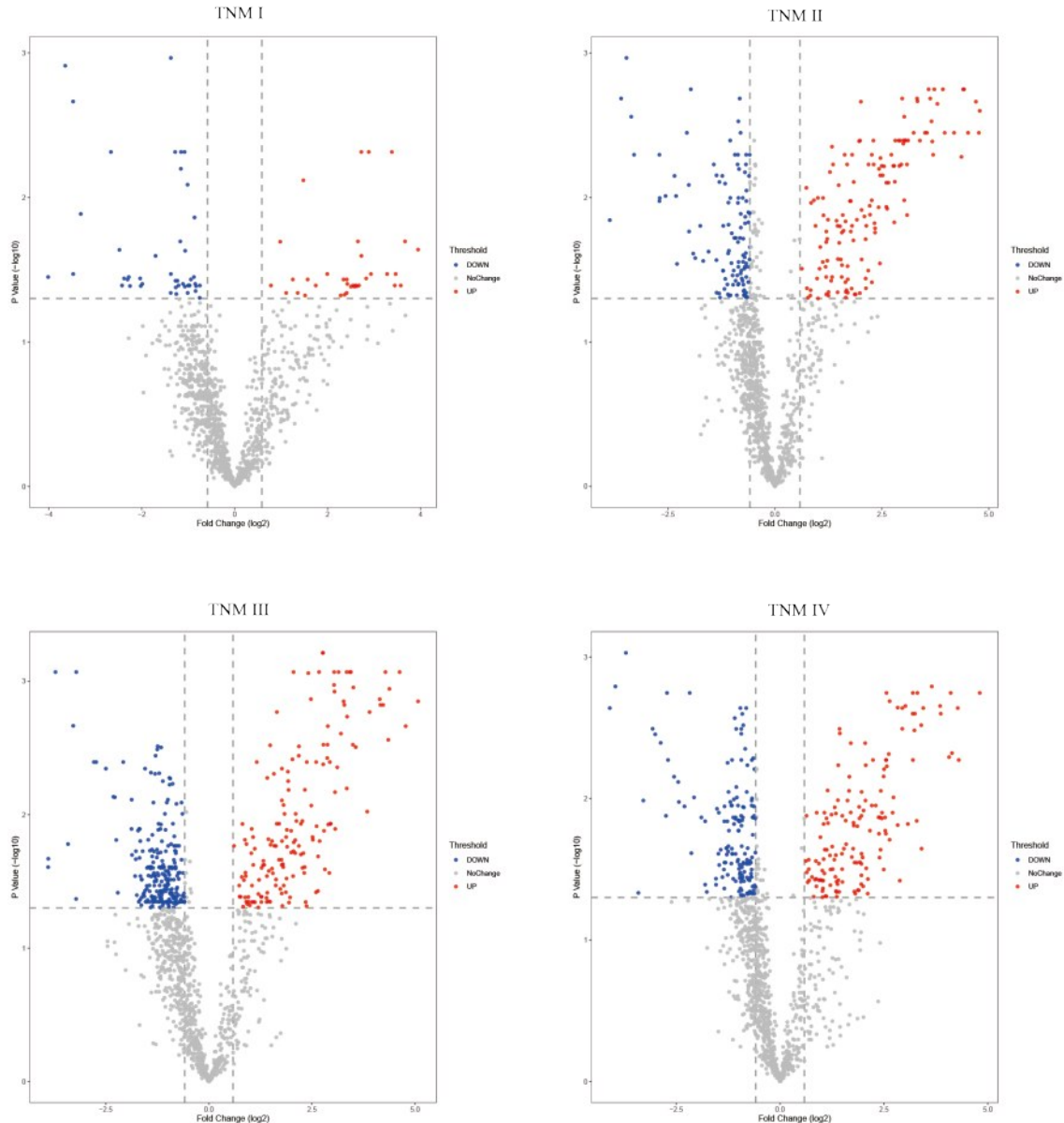


图3 不同TNM时期CRC患者与健康人相比血浆差异表达蛋白火山图

Fig. 3 Volcano plots of differentially expressed proteins comparing CRC patients at different TNM stages with healthy controls

将各个TNM时期的差异表达蛋白进行韦恩图交集分析, 筛选得到在本研究样本的全部TNM分期对比健康对照时, 均达到上述差异筛选阈值的蛋白, 共得到上调蛋白32个, 下调蛋白19个(图4), 该组蛋白为多组比较得到的交集统计集合, 上/下调表达蛋白信息详见附表3。通过对其进行文献调研, 发现其中多个代表性分子已在临床队列中被证实与肿瘤浸润、分期、转移、预后及化疗响应等临床指标显著相关, 具备多分期无创血浆标志物潜力^[21]。糖酵解关键酶ENO1在各分期CRC血浆持续上调, 其循环表达水平与肿瘤浸润深度、淋巴结转移正相关, 可通过激活PI3K/Akt通路驱动糖代谢重编程与肿瘤侵袭, 是区分进展期CRC与健康人群的独立血浆标志物^[22, 23]; 细胞外囊泡蛋白SPARC作为跨分期稳定上调分子, 对结肠癌具备良好诊断效能, 丰度与原发灶部位、术后复发风险相关, 同时可调控CRC对5-氟尿嘧啶药物敏感性^[24, 25]; 四次跨膜蛋白CD9富集于CRC患者循环细胞外囊泡, 血浆EV-CD9表达量可梯度区分腺瘤、I-IV期CRC与健康对照, 多中心队列证实其可作为CRC无进展生存期独立预测因子, 是液体活检经典囊泡标志物^[26]; 血浆蛋白质组队列证实APOA4在I-IV期全部CRC患者血浆中一致性下调, 跨分期样本对比显示APOA4降低伴随全身炎症标志物升高, 联合LRG1可有效区分CRC人群与健康对照^[27, 28]; 补体组分C1QC主要由肿瘤相关巨噬细胞分泌, 肿瘤组织高C1QC浸润对应免疫抑制微环境与免疫检查点抑制剂疗效不佳, 外周循环补体组分改变可反映贯穿CRC全病

程的系统性免疫失衡^[29]；血浆糖蛋白质组学证实 FN1 在各分期 CRC 血浆中一致性下调，其糖基化修饰谱存在分期特异性差异，循环 FN1 降低与细胞外基质屏障破坏、肿瘤侵袭转移能力增强直接相关^[30]；剩余交集蛋白虽多数具备调控 CRC 增殖、迁移、代谢的细胞与组织实验证据，但缺乏覆盖全部 TNM 分期的血浆临床验证队列，上述代表性全程紊乱蛋白共同构成覆盖 CRC 全病程、兼具诊断与预后价值的血浆蛋白特征谱，为多分子联合液体活检标志物开发提供可靠候选集。

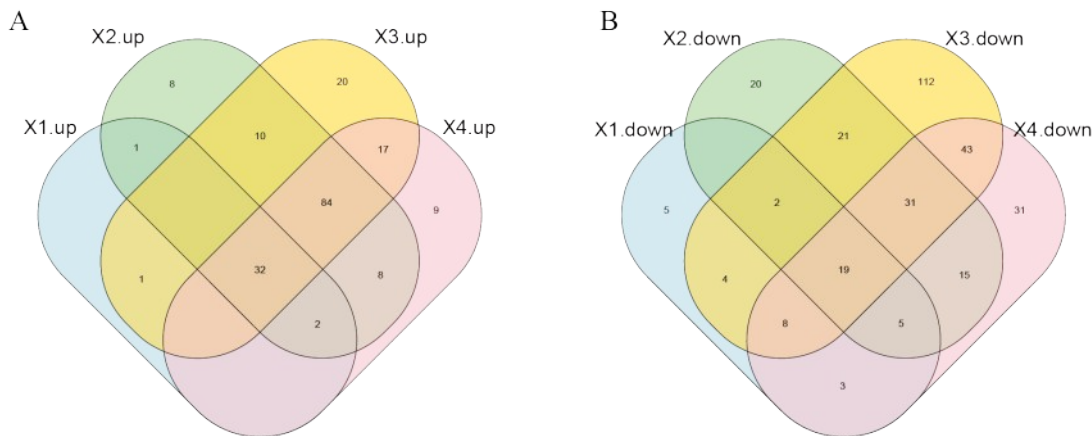


图4 不同时期CRC患者上调(A)和下调(B)表达蛋白比较韦恩图

Fig. 4 Venn diagrams for comparison of upregulated (A) and downregulated (B) proteins in CRC patients at different stages

2.4 结直肠癌血浆中的差异表达蛋白功能富集分析

为进一步解析本研究筛选得到的交集蛋白的功能特征，对该部分蛋白开展GO富集分析，涵盖生物过程(Biological process, BP)、细胞组分(Cellular component, CC)和分子功能(Molecular function, MF)3个维度，并结合KEGG通路富集分析，探究该蛋白集合所关联的信号通路。

对本队列各TNM分期对比健康对照均达到差异阈值的上调蛋白进行富集分析，结果提示该组蛋白富集条目主要涉及血小板活化、细胞-基质黏附、肌动蛋白细胞骨架调控、氧化应激应答以及碳水化合物代谢等生物学过程(图5)。血浆样本中富集到大量血小板相关条目，提示结直肠癌患者外周循环内血小板相关蛋白水平发生改变，反映出循环凝血系统的扰动；细胞黏附、整合素信号、细胞骨架相关通路的富集，说明循环中与细胞迁移、细胞间互作相关的蛋白丰度出现变化；凋亡负调控、氧化应激应答、糖异生-糖酵解代谢相关条目富集，则提示机体循环系统中应激调控与能量代谢相关蛋白谱发生改变。

本队列交集筛选得到的下调蛋白，富集条目主要涉及免疫防御、组织修复、稳态调控及基础代谢等生理过程(图6)。免疫应答、补体激活、抗原结合相关功能出现富集下调，提示患者外周循环中参与固有免疫调控的蛋白丰度发生改变，与既往报道中肿瘤环境下免疫监视相关分子的变化特征相呼应。囊泡转运、伤口应答相关条目富集水平降低，这类蛋白参与肠道黏膜损伤后的修复过程，该富集趋势可反映循环中参与组织修复、肠道屏障稳态调控的蛋白谱发生改变。细胞骨架组分、蛋白水解活性、脂类与维生素代谢通路的下调，则提示机体循环微环境稳态以及全身物质代谢相关蛋白出现扰动。

2.5 结直肠癌血浆蛋白疾病标志物筛选

将各分期CRC患者血浆蛋白分别与健康人群进行差异分析，结合多组差异蛋白集的韦恩图交叉筛选，最终鉴定出9个仅在IV期CRC患者血浆中特异性高表达的蛋白，包括HSPB1、CSTB、TIMP1、SI00P、GDI2、FABP4、ANXA3、PSMB3和MASP1。气泡图结果显示(图7A)，相较于健康对照，该9个蛋白在IV期上调最为显著，在I-III期可见一定上调趋势，但上调幅度及统计学显著性不及IV期，提示其与CRC晚期恶性进展密切相关，有望成为用于晚期结直肠癌识别的候选特征分子，值得后续开展更大样本队列进一步评价。

为初步探讨本研究筛选得到的血浆差异蛋白其编码基因在肿瘤组织内部的转录状态，我们借助TCGA结直肠癌转录组数据集开展辅助分析。利用TCGA数据库样本，对上述9个蛋白对应的基因开展主成分分析。结果显示，正常结肠组织(Colon-Sigmoid_GTEX)、癌旁组织(COAD_Peritumor)与结肠

癌组织 (COAD_Tumor) 在 PCA 散点图中各自独立聚类, PC1 (74.7%) 将三类样本区分度良好 (图 7B)。综上, 该 9 基因分子组合大概率参与结肠癌晚期的恶性演进过程, 具备作为结肠癌良恶性鉴别、临床分期判定及晚期病变预警候选标志物的潜力。

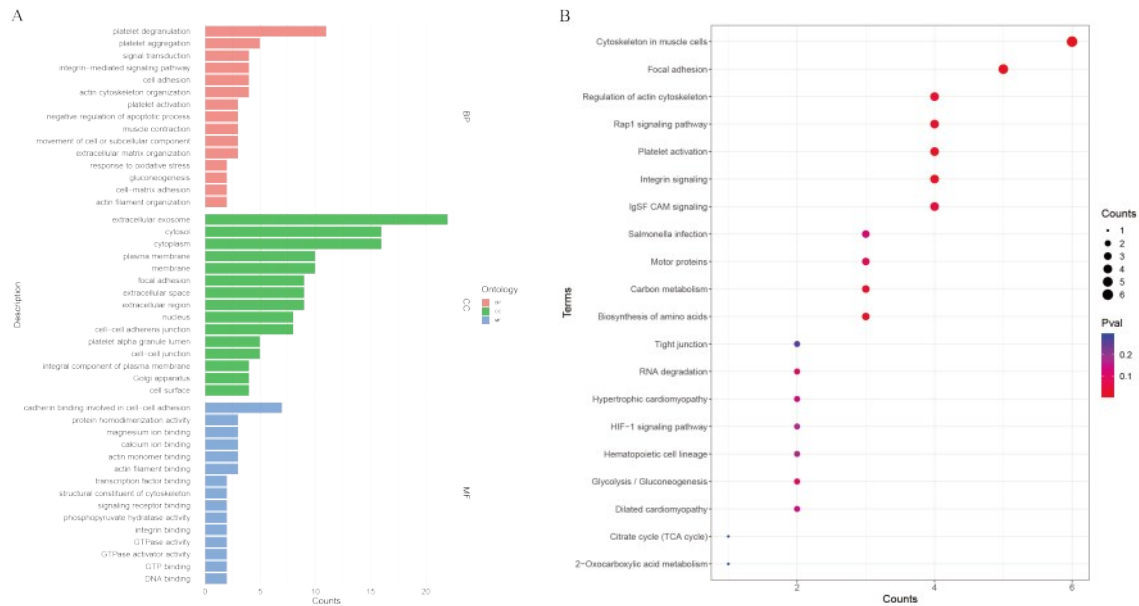


图 5 CRC 患者上调表达蛋白 GO 富集分析条形图 (A) 和 KEGG 富集分析气泡图 (B)

Fig. 5 Bar plot of GO analysis (A) and bubble plot of KEGG enrichment analysis (B) for upregulated proteins in CRC patients

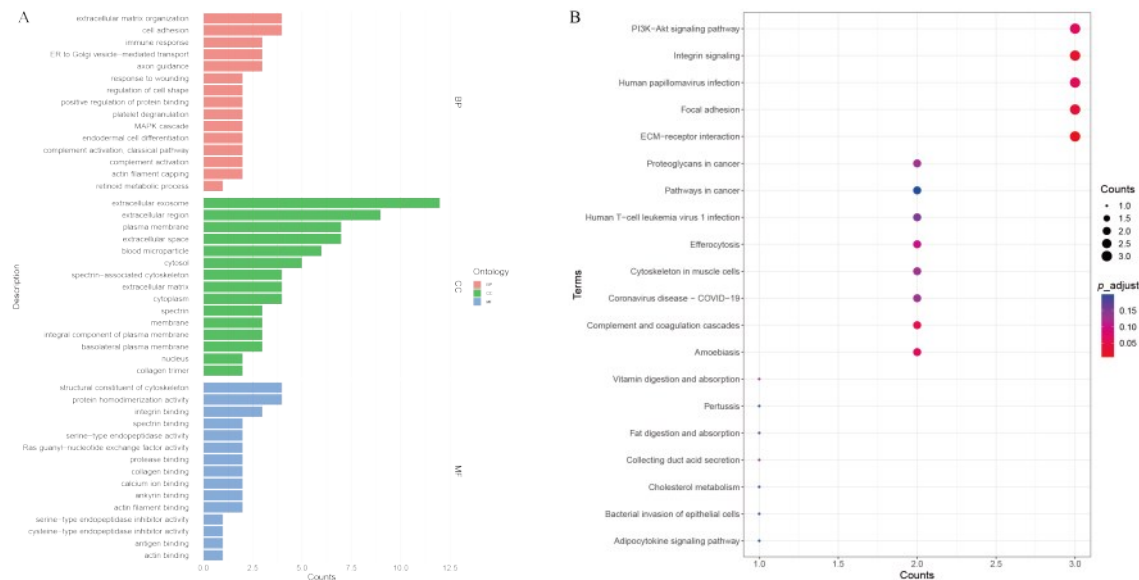


图 6 CRC 患者下调表达蛋白 GO 富集分析条形图 (A) 和 KEGG 富集分析气泡图 (B)

Fig. 6 Bar plot of GO analysis (A) and bubble plot of KEGG enrichment analysis (B) for downregulated proteins in CRC patients

在验证组血浆样本中 FABP4、MASP1、TIMP1 蛋白同样呈现出 CRC 患者组高表达且在 IV 期上调幅度最为突出的趋势 (图 8A)。对候选分子 TIMP1 进行 TCGA 数据库表达分析, 结果显示, 在结直肠癌 (Colon Adenocarcinoma, COAD 和 Rectum Adenocarcinoma, READ) 样本中, 肿瘤组织内 TIMP1 的表达水平显著高于正常组织 (图 8B)。进一步生存分析表明, 以 TIMP1 表达量为界将患者分为高表达组与低表达组后, 两组总体生存期存在明显差异 (Logrank $p=0.02$); TIMP1 高表达患者的死亡风险显著上升, 整体预后更差 (图 8C)。上述结果证实, TIMP1 在 CRC 中异常上调, 且其高表达与患者不良预后密切相关。

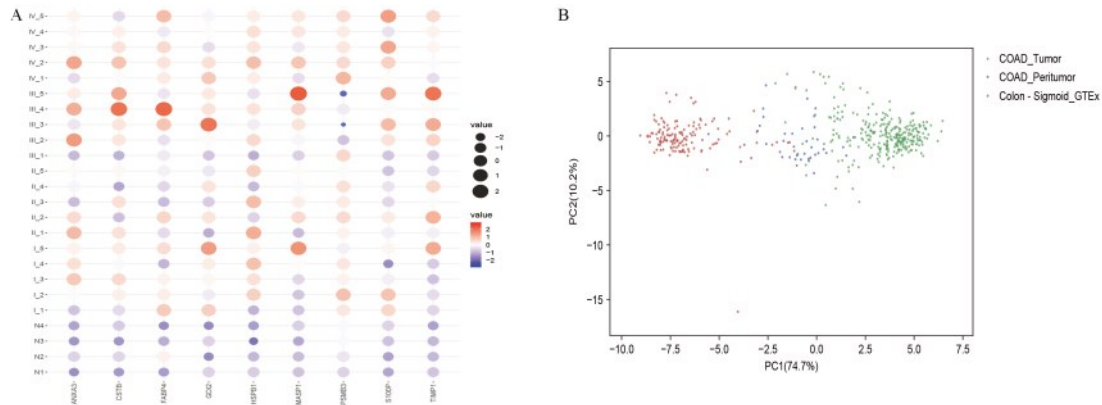


图7 IV期CRC患者血浆优势高表达蛋白气泡图(A)及其编码基因在TCGA结肠组织样本的主成分分析散点图(B)
Fig. 7 Bubble plot of predominantly highly expressed plasma proteins in stage IV CRC patients (A) and PCA scatter plot of their encoding genes in TCGA colon tissue samples (B)

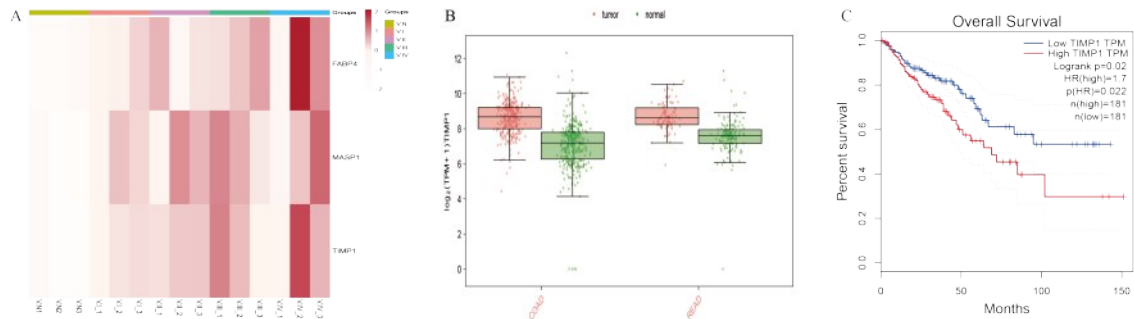


图8 关键候选分子在验证组血浆中的表达情况(A),TIMP1在TCGA数据库CRC队列中的表达箱线图(B)及TIMP1的CRC患者生存分析(C)
Fig. 8 Expression of key candidate molecules in plasma from the validation cohort (A), box plot of TIMP1 in the CRC cohort from the TCGA database (A) and survival analysis of TIMP1 in CRC patients (B)

3 结 论

本研究纳入不同TNM分期的CRC血浆队列,联合高丰度蛋白去除与directDIA定量蛋白质组技术完成质谱检测,共鉴定1736个蛋白群组,可检出CEA、MMP9等经典肿瘤相关蛋白。经多组比较交集筛选得到本队列内32个上调、19个下调蛋白,ENO1、SPARC、CD9等分子在本研究各TNM分期相对健康对照均达到差异筛选阈值,可作为广谱无创诊断的候选分子。功能富集分析结果显示,上调蛋白富集于血小板活化、糖酵解、细胞侵袭相关条目;下调蛋白集中于固有免疫、补体激活及黏膜修复相关生物学过程,反映了结直肠癌进展过程中循环层面的蛋白谱变化特征。本研究筛选获得9个IV期优势高表达候选蛋白,结合TCGA公共数据集开展辅助分析,该组基因在正常、癌旁及肿瘤组织中呈现出不同的表达分布特征。候选分子TIMP1高表达与患者不良预后存在关联。综上,本研究解析了本队列中结直肠癌不同分期的血浆蛋白谱变化,得到与疾病进展相关的紊乱通路信息以及广谱、晚期相关候选蛋白集合,可为结直肠癌液体活检相关研究提供分子参考。本研究样本规模有限,存在统计偏倚风险,后续需扩大大中心队列验证标志物诊断效能,推动其临床转化。

参考文献:

- [1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA: a cancer journal for clinicians*, 2024, 74 (3): 229-63.
- [2] NGUYEN L H, GOEL A, CHUNG D C. Pathways of Colorectal Carcinogenesis [J]. *Gastroenterology*, 2020, 158 (2): 291-302.
- [3] BENSON A B, VENOOK A P, ADAM M, et al. Colon Cancer, Version 3.2024, NCCN Clinical Practice Guide-

- lines in Oncology [J]. *Journal of the National Comprehensive Cancer Network : JNCCN*, **2024**, 22 (2 d).
- [4] SIEGEL R L, KRATZER T B, GIAQUINTO A N, et al. *Cancer statistics*, 2025 [J]. CA: a cancer journal for clinicians, **2025**, 75 (1): 10–45.
- [5] URBIOLA-SALVADOR V, JABŁOŃSKA A, MIROSZEWSKA D, et al. Plasma protein changes reflect colorectal cancer development and associated inflammation [J]. *Front Oncol*, **2023**, 13: 1158261.
- [6] MAGNUSSON M I, AGNARSSON B A, JONASSON J G, et al. Histopathology and levels of proteins in plasma associate with survival after colorectal cancer diagnosis [J]. *British journal of cancer*, **2023**, 129 (7): 1142–51.
- [7] VAN EYK J E, SNYDER M P. Precision Medicine: Role of Proteomics in Changing Clinical Management and Care [J]. *Journal of proteome research*, **2019**, 18 (1): 1–6.
- [8] AEBERSOLD R, MANN M. Mass-spectrometric exploration of proteome structure and function [J]. *Nature*, **2016**, 537 (7620): 347–55.
- [9] LUDWIG C, GILLET L, ROSENBERGER G, et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial [J]. *Molecular systems biology*, **2018**, 14 (8): e8126.
- [10] GEYER P E, HOLDT L M, TEUPSER D, et al. Revisiting biomarker discovery by plasma proteomics [J]. *Molecular systems biology*, **2017**, 13 (9): 942.
- [11] SMIT E R, VAN DER ZWAAN C, GROTEN S A, et al. Considering Measurement Time and Depth in diaPASEF Plasma Proteomics [J]. *Journal of proteome research*, **2025**, 24 (11): 5904–10.
- [12] SHISHKOVA E, COON J J. Rapid preparation of human blood plasma for bottom-up proteomics analysis [J]. *STAR protocols*, **2021**, 2 (4): 100856.
- [13] PANG C, XU F, LIN Y, et al. LC-MS/MS analysis reveals plasma protein signatures associated with lymph node metastasis in colorectal cancer [J]. *Frontiers in immunology*, **2024**, 15: 1465374.
- [14] PATEL S K, BONIS J, ROSE J P, et al. Exosomes released from senescent cells and circulatory exosomes isolated from human plasma reveal aging-associated proteomic and lipid signatures [J]. *Aging*, **2025**, 17 (8): 1929–65.
- [15] NATARAJAN R A, KALIYAMOORTHY S, BOOPATHI V, et al. Mapping the stage-specific plasma p53 interactome reveals colorectal cancer progression signatures and therapeutic vulnerabilities [J]. *The journal of liquid biopsy*, **2025**, 10: 100329.
- [16] HALL C, CLARKE L, PAL A, et al. A Review of the Role of Carcinoembryonic Antigen in Clinical Practice [J]. *Annals of coloproctology*, **2019**, 35 (6): 294–305.
- [17] BUTTACAVOLI M, DI CARA G, ROZ E, et al. Integrated Multi-Omics Investigations of Metalloproteinases in Colon Cancer: Focus on MMP2 and MMP9 [J]. *International journal of molecular sciences*, **2021**, 22 (22).
- [18] ZHOU J, JI Q, LI Q. Resistance to anti-EGFR therapies in metastatic colorectal cancer: underlying mechanisms and reversal strategies [J]. *Journal of experimental & clinical cancer research: CR*, **2021**, 40 (1): 328.
- [19] YANQING H, CHENG D, LING X. Serum CA72-4 as a Biomarker in the Diagnosis of Colorectal Cancer: A Meta-analysis [J]. *Open medicine (Warsaw, Poland)*, **2018**, 13: 164–71.
- [20] HALL N R, FINAN P J, STEPHENSON B M, et al. The role of CA-242 and CEA in surveillance following curative resection for colorectal cancer [J]. *British journal of cancer*, **1994**, 70 (3): 549–53.
- [21] LI Y, WANG B, YANG W T, et al. Longitudinal plasma proteome profiling reveals the diversity of biomarkers for diagnosis and cetuximab therapy response of colorectal cancer [J]. *Nat Commun*, **2024**, 15 (1).
- [22] GU J R, ZHONG K Q, WANG L G, et al. ENO1 contributes to 5-fluorouracil resistance in colorectal cancer cells via EMT pathway [J]. *Front Oncol*, **2022**, 12.
- [23] LIU Y, HOU Y Y, ZHANG F, et al. ENO1 deletion potentiates ferroptosis and decreases glycolysis in colorectal cancer cells via AKT/STAT3 signaling [J]. *Exp Ther Med*, **2024**, 27 (4).
- [24] ZHONG M E, CHEN Y Y, XIAO Y, et al. Serum extracellular vesicles contain SPARC and LRG1 as biomarkers of colon cancer and differ by tumour primary location [J]. *Ebiomedicine*, **2019**, 50: 211–23.
- [25] TAI I T, DAI M, OWEN D A, et al. Genome-wide expression analysis of therapy-resistant tumors reveals SPARC as a novel target for cancer therapy [J]. *J Clin Invest*, **2005**, 115 (6): 1492–502.
- [26] WEI P, WU F, KANG B, et al. Plasma extracellular vesicles detected by Single Molecule array technology as a liquid biopsy for colorectal cancer [J]. *J Extracell Vesicles*, **2020**, 9 (1).
- [27] ZHU W Y, LI M Z, WANG Q S, et al. Quantitative Proteomic Analysis Reveals Functional Alterations of the Peripheral Immune System in Colorectal Cancer [J]. *Mol Cell Proteomics*, **2024**, 23 (6).
- [28] AHN S B, SHARMA S, MOHAMEDALI A, et al. Potential early clinical stage colorectal cancer diagnosis using a proteomics blood test panel [J]. *Clin Proteom*, **2019**, 16 (1).
- [29] ZHANG H, HONG L, DONG Z, et al. Spatially resolved C1QC (+) macrophage-CD4 (+) T cell niche in colorectal cancer microenvironment: implications for immunotherapy response [J]. *Cell Discov*, **2025**, 11 (1): 60.
- [30] CHANTARAAMPORN J, CHAMPATTANACHAI V, KHONGMANEE A, et al. Glycoproteomic Analysis Reveals Aberrant Expression of Complement C9 and Fibronectin in the Plasma of Patients with Colorectal Cancer [J]. *Proteomes*, **2020**, 8 (3).

(责任编辑：丁 岩)