



基于蛋白质组学技术的新型冠状病毒肺炎研究进展*

刘响^{1,2)} 宫鹏云^{1,2)} 汤敏^{1,2)} 胡鸿珂^{1,2)} 张永彪^{1,2)**} 刘超^{1,2)**}

(¹⁾北京航空航天大学医学科学与工程学院, 北京 100191; (²⁾北京航空航天大学生物与医学工程学院, 北京 100191)

摘要 新型冠状病毒肺炎 (coronavirus disease 2019, COVID-19) 是一种由严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) 引发的传染病。此种病毒传染性强、传播速度快, 对全球人民的身体健康和生命安全造成严重威胁。蛋白质组学技术以其高通量、高灵敏度的特点, 在疾病生物标志物的发现、分子机制研究及治疗靶点研究中扮演着重要角色, 并被广泛应用于 COVID-19 的研究中。本文介绍了 SARS-CoV-2 的基因组结构及病毒感染过程, 总结了目前常用的基于质谱的蛋白质组学研究技术, 重点综述了蛋白质组学技术在 COVID-19 生物标志物的发现、分子机制研究和药物治疗靶标研究中的应用进展, 最后展望了蛋白质组学的未来发展方向, 以期能够有助于推动蛋白质组学技术在 COVID-19 精准诊断和治疗中的发展。

关键词 蛋白质组学, 质谱, 新型冠状病毒肺炎, 生物标志物

中图分类号 Q51, R34

DOI: 10.16476/j.pibb.2022.0029

2019年12月, 一种新型冠状病毒 (后命名为严重急性呼吸综合征冠状病毒 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2)) 引起的肺炎疫情爆发, 世界卫生组织 (World Health Organization, WHO) 将该疾病命名为新型冠状病毒肺炎 (coronavirus disease 2019, COVID-19), 简称“新冠肺炎”, 并于2020年3月宣布 COVID-19 成为全球性大流行病^[1]。截至2022年3月18日, 全球已累计报告超4.6亿名确诊病例, 逾608万名患者死亡, 目前仍在持续扩散中^[2]。COVID-19 现在是影响全球公共健康的最具破坏性的流行病之一, 引发国内外广泛关注。为了找到遏制 COVID-19 大流行的解决方案, 全球研究工作已迅速动员起来, 在基础研究、诊断、治疗和药物研发等方面开展了一系列研究并取得了重要进展。

蛋白质组学 (proteomics) 是1994年由 Marc R. Wikins 首先提出, 是以所有基因表达的全部蛋白质为研究对象, 研究细胞、组织或生物体蛋白质组成及其变化规律的科学^[3-4]。蛋白质组学技术是当前解决医学和生物学中诸多问题的重要工具之一, 已被广泛应用于传染性疾病的研究当中。通过对病

原体和感染后的宿主细胞、组织、体液等进行蛋白质组学研究, 同时结合免疫学进行分析, 可为传染性疾病的诊断、治疗、预防等方面的研究提供新的策略和依据。COVID-19 疫情爆发以来, 蛋白质组学技术已被广泛应用于 COVID-19 的研究中, 并在筛选生物标志物、研究病毒感染的分子机制和研发预防和治疗药物等3个方面取得较大进展。

本文首先介绍了 SARS-CoV-2 的基因组结构和病毒感染过程, 总结了目前常用的基于质谱的蛋白质组学研究技术, 并对基于质谱的蛋白质组学技术在 COVID-19 生物标志物研究、感染机制研究和药物治疗靶标研究中的应用进展进行了梳理、总结和展望, 以期对 COVID-19 的防控和治疗提供有力的支撑。

* 国家自然科学基金 (32171442) 和科技部国家重点研发计划 (2021YFA1301602, 2021YFA1301603) 资助项目。

** 通讯联系人。

刘超 Tel: 010-82338500, E-mail: liuchaobuaa@buaa.edu.cn

张永彪 Tel: 010-82338500, E-mail: zhangyongbiao@buaa.edu.cn

收稿日期: 2022-01-21, 接受日期: 2022-04-19

1 SARS-CoV-2的基因组结构和病毒感染过程

COVID-19是由SARS-CoV-2引起的疾病^[5]。经基因组测序, SARS-CoV-2为有包膜的单股正链RNA病毒, 其基因组序列与严重急性呼吸综合征冠状病毒(severe acute respiratory syndrome coronavirus, SARS-CoV)序列相似度约为80%, 与中东呼吸综合征冠状病毒(Middle East respiratory syndrome coronavirus, MERS-CoV)序列相似度约为50%^[6-7], 属于 β 冠状病毒属, 是一种新型的冠状病毒^[5, 8]。

SARS-CoV-2的基因组大小约29 900个碱基,

包含14个主要的开放阅读框(open reading frame, ORF), 编码27种蛋白质(图1), 位于基因组5'端的ORF1ab和ORF1a基因分别编码2个多聚蛋白pp1ab和pp1a, 多聚蛋白pp1ab和pp1a经剪切可产生16种非结构蛋白(non-structural proteins, NSPs), 包括nsp1~16。基因组的3'端编码4种主要的结构蛋白和8种辅助蛋白, 其中刺突蛋白(spike protein, S蛋白)、包膜蛋白(envelope protein, E蛋白)、膜蛋白(membrane glycoprotein, M蛋白)、核衣壳蛋白(nucleocapsid protein, N蛋白)是结构蛋白; ORF3a、3b、p6、7a、7b、8b、9b和ORF14是辅助蛋白^[7, 9]。这些蛋白质是SARS-CoV-2感染宿主的物质基础。

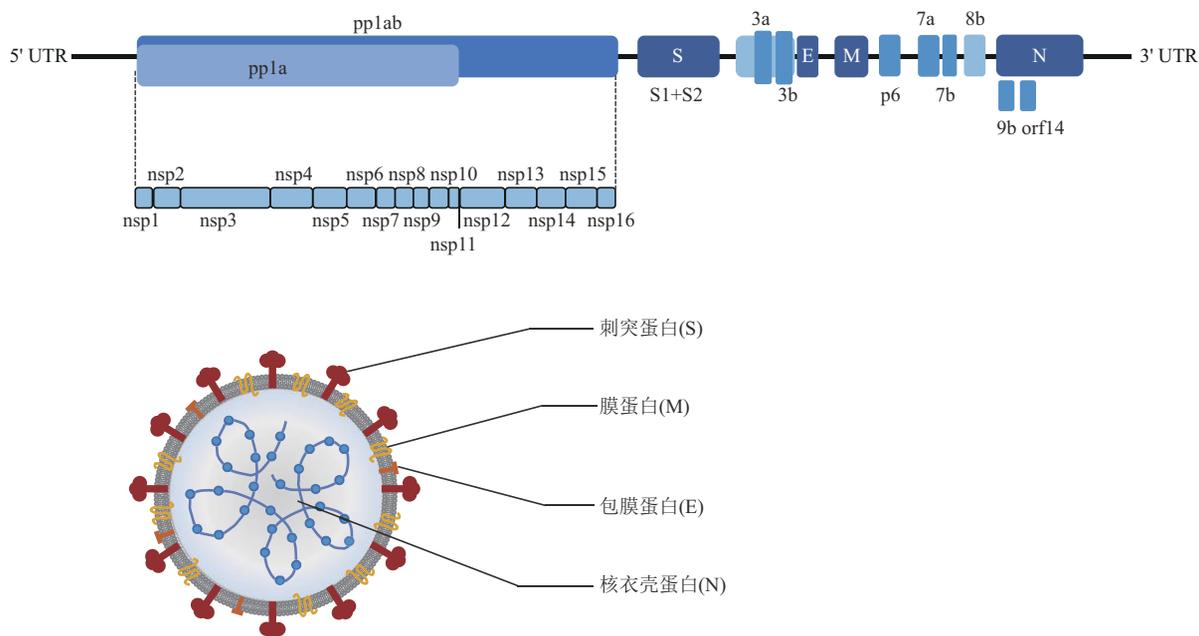


Fig. 1 Genome composition of SARS-CoV-2 and diagram of structure of SARS-CoV-2 particle

图1 SARS-CoV-2病毒基因组组成和主要蛋白质分布及结构示意图

SARS-CoV-2的基因组大小约29 900个碱基, ORF1ab和ORF1a基因位于基因组的5'端, 分别编码2个多聚蛋白pp1ab和pp1a, 多聚蛋白pp1ab和pp1a经剪切可产生16种非结构蛋白(nsp1至nsp16), 基因组的3'端包含4个结构蛋白基因(S、E、M、N)和8个附属蛋白基因(3a、3b、p6、7a、7b、8b、9b、orf14)。

SARS-CoV-2感染宿主细胞是通过病毒表面的S蛋白与宿主细胞表面的跨膜蛋白血管紧张素转化酶2(angiotensin converting enzyme 2, ACE2)发生相互作用所介导的^[7, 10]。S蛋白由2个糖基化的亚基S1和S2组成, 其中S1亚基含有受体结合域(receptor binding domain, RBD), 负责识别细胞的受体; S2亚基含有膜融合过程所需的基本元件,

催化病毒和细胞膜的融合^[10]。ACE2是一种细胞表面肽酶, 可水解血管紧张素II, ACE2是目前已知最主要的SARS-CoV-2受体, 在大多数器官中均有表达, 特别是在肺和小肠上皮中高表达^[11]。病毒感染宿主细胞的过程(图2)包括: ①病毒的S蛋白与宿主细胞表面受体结合后, 通过两种方式进入细胞, 一种是被质膜上的跨膜丝氨酸蛋白酶2

(TMPRSS2) 切割S蛋白，激活病毒膜与宿主膜融合，另一种是被组织蛋白酶L (CTSL) 切割和激活S蛋白，介导 SARS-CoV-2 通过内吞进入细胞，基因组RNA 释放到细胞质中。②宿主核糖体结合到病毒RNA，翻译pp1a、pp1ab多聚蛋白。③产生的多聚蛋白pp1a和pp1ab被翻译后加工成单个非结构蛋白 (nsp1~16)。④RNA 依赖 RNA 聚合酶 (RNA-dependent RNA polymerase, RdRp) 与 nsp7、nsp8 等因子形成转录复制复合体 (replication and transcription complex, RTC) 复制病毒RNA，负链RNA 可用于翻译各种结构蛋白，

如E、N、S和M蛋白。⑤复制后的RNA由N蛋白包裹与M蛋白、E蛋白和S蛋白组装成完整的病毒颗粒。⑥病毒通过高尔基体或溶酶体释出感染其他细胞 [12-15]。

ACE2广泛分布于人体的多个组织，包括肺、小肠、心脏、肝脏、肾脏、胰腺、睾丸和神经组织等 [15-19]。因此SARS-CoV-2能够侵染人体的多个组织，对人体造成严重伤害，导致严重的呼吸系统问题、免疫功能障碍、神经系统损伤以及心血管、肺、胃肠道、肾脏功能障碍和多器官衰竭等 [16-22]。

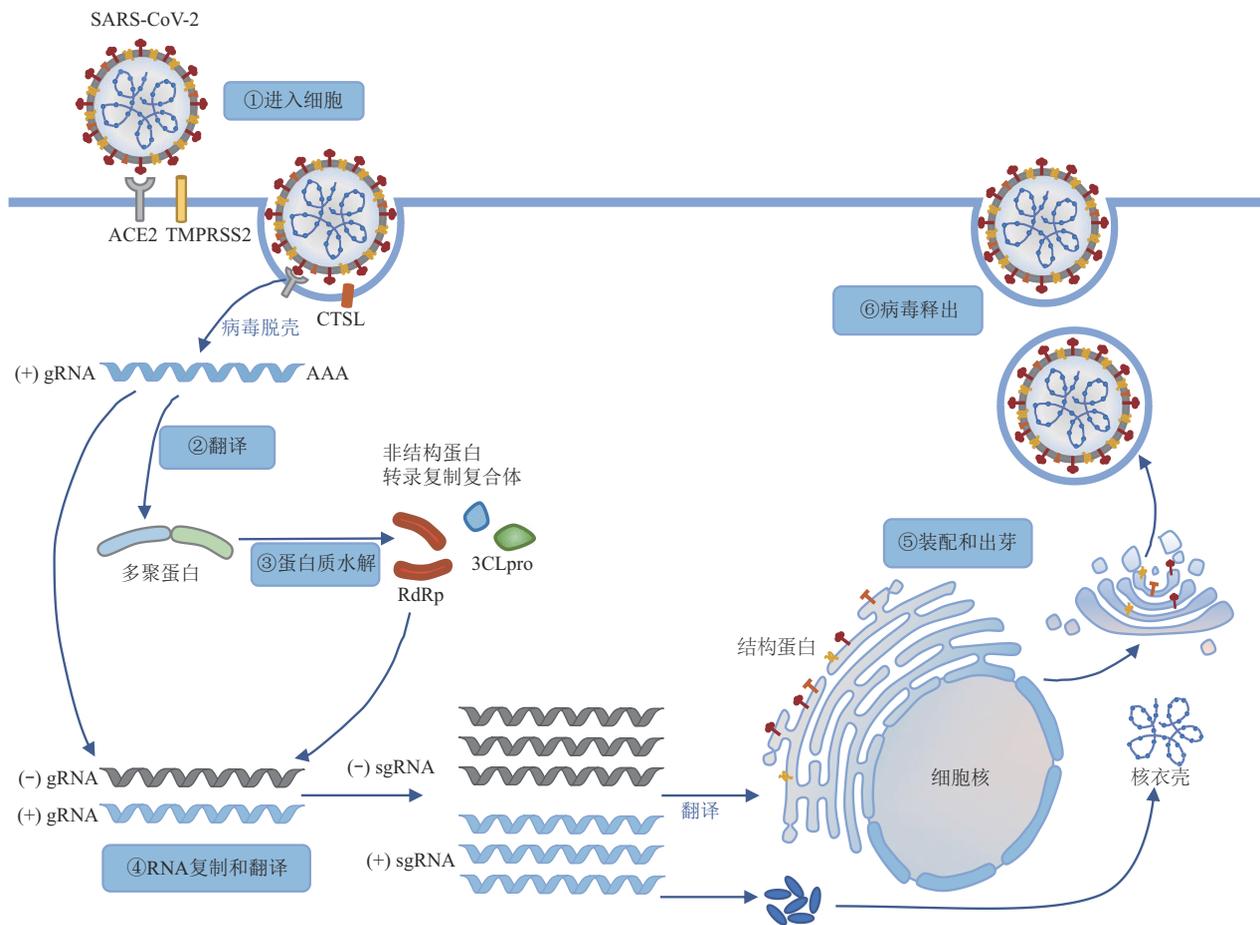


Fig. 2 The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lifecycle

图2 SARS-CoV-2的生命周期

ACE2: 血管紧张素转化酶2 (angiotensin converting enzyme 2); TMPRSS2: 跨膜丝氨酸蛋白酶2 (transmembrane protease, serine 2); CTSL: 组织蛋白酶L (cathepsin L); RdRp: RNA依赖RNA聚合酶 (RNA-dependent RNA polymerase); 3CLpro: 3C样蛋白酶 (3C-like protease); gRNA: 基因组RNA (genomic RNA); sgRNA: 亚基因组RNA (subgenomic RNA)。

2 基于质谱的蛋白质组学技术

生物质谱技术是蛋白质组学研究中最重要分析技术。目前常用的质谱包括两种: 基质辅助激光解吸电离-飞行时间质谱 (matrix-assisted laser desorption ionization time-of-flight mass spectrometry, MALDI-TOF MS) 和液相色谱-电喷雾质谱联用 (liquid chromatography-electrospray ionization mass spectrometry, LC-ESI-MS)^[23]。两者都可以用于蛋白质分析鉴定, 但是电离原理不同, 应用方向各有特点。MALDI-TOF MS 常用于简单样品如纯化蛋白质的分析鉴定, 其特点是速度快、灵敏度高, 常用于蛋白质、多肽、核酸等的分析检测, 以及微生物菌种鉴定和质谱成像等, 在 SARS-CoV-2 病毒颗粒的鉴定和生物标志物研究中发挥重要作用; LC-ESI-MS 主要用于来自组织、细胞等复杂蛋白质与多肽样品的分析检测, 其特点是灵敏度高、检测范围广, 在 COVID-19 的生物标志物、发病机制和药物靶标研究中都发挥了重要的作用。

2.1 MALDI-TOF MS 技术

MALDI-TOF MS 技术的原理是将蛋白质、多肽等样品溶液与芳香族有机酸等小分子基质溶液混合后点在金属靶盘上, 待溶剂挥发后形成共结晶, 用激光照射晶体时, 基质分子吸收激光能量、蒸发和离子化, 并带着样品分子解吸附进入气相, 基质-样品分子之间发生电荷转移使样品分子电离, 采用飞行时间质量分析器检测带电荷的分析物, 从而获得样品的质荷比 (m/z) 信息。MALDI-TOF MS 一般用于蛋白质、多肽、核酸、脂质和寡糖的分子质量测定、蛋白质的肽质量指纹图谱分析与快速鉴定。MALDI-TOF/TOF MS 可得到 MS/MS 谱图, 可以得到肽段氨基酸组成信息。

MALDI-TOF MS 常用于微生物菌种快速鉴定。每一种微生物都有其特征的蛋白质指纹图谱, 通过 MALDI-TOF MS 分析获得微生物菌种核糖体蛋白指纹图谱, 与已知微生物菌种蛋白质指纹图谱数据库比对分析后, 实现微生物菌种的快速鉴定。MALDI-TOF MS 数据库中菌株种类丰富, 来源广泛, 在临床上常用于常见致病菌检测。基于 MALDI-TOF MS 的蛋白质组学及多肽组学技术也成为 SARS-CoV-2 快速检测的潜在技术, 已有研究报道利用 MALDI-TOF MS 技术鉴定鼻咽拭子中的 SARS-CoV-2 病毒蛋白, 检测准确度可达到

93.9%^[24-25], 这些研究揭示了 MALDI-TOF MS 技术的临床诊断应用价值。

2.2 LC-ESI-MS 技术

目前蛋白质组学研究主要使用 LC-ESI-MS 联用技术, 其原理是: 样品经色谱柱分离后, 经毛细管喷头流出时, 在千伏级高压电场的作用下形成带电小液滴; 随着液滴的溶剂蒸发, 液滴表面离子密度增大, 当库伦斥力和液滴表面张力极限值相等时发生库伦爆炸, 形成更小的带电液滴; 然后在电场作用下重复蒸发、分裂的过程, 直至形成气相离子进入质量分析器被分离检测。该方法的优点是可以实现从液态到气态分子的转变, 产生的离子可以带有一个或多个电荷, 这种多电荷离子的产生大大扩展了普通质谱仪能分析的质量范围, 使质谱仪可以分析分子质量为几十万道尔顿的蛋白质分子^[23, 26]。因此, LC-ESI-MS 技术在蛋白质组学研究中应用最为广泛, 适用于蛋白质质谱鉴定、蛋白质组定量分析、蛋白质翻译后修饰鉴定和蛋白质相互作用研究等。

目前基于 LC-ESI-MS 的蛋白质组学技术主要采用 3 种采集方式: 数据依赖性采集 (data dependent acquisition, DDA)、数据非依赖性采集 (data independent acquisition, DIA) 和靶向数据采集^[27]。

2.2.1 数据依赖性采集 (DDA) 技术

DDA 技术的原理是利用一级全扫描检测肽段母离子, 然后按照母离子的信号强度排序, 将强度较高的一系列 (topN) 母离子由强到弱依次进行选择、碎裂和检测二级碎片离子。同时, 质谱的动态排除、动态背景扣除、价态排除等技术, 使 DDA 尽可能多地采集有效母离子的二级谱图, 实现鉴定结果最大化^[28]。DDA 技术在 SARS-CoV-2 感染后的病毒蛋白质鉴定、蛋白质组定量、蛋白质翻译后修饰 (如 S 蛋白的糖基化以及蛋白质组的磷酸化等) 鉴定和蛋白质相互作用研究中发挥重要作用。基于 DDA 的同位素标记定量是目前应用最广泛的蛋白质组学定量技术, 其中使用最多的同位素标记试剂是 iTRAQ (isobaric tags for relative and absolute quantitation) 和 TMT (tandem mass tags)。标记定量的主要策略是向每组多肽样品中引入不同的稳定同位素标记的小分子, 通过同位素标记后所产生的质量差来识别肽段的来源。标记定量方法将不同组样本混匀后同时进行质谱检测, 可以避免样品前处理所带来的定量误差^[29]。目前标记定量最

多可实现16标同时定量,特别适用于采用多种处理方式或来自多个处理时间的COVID-19样本的差异蛋白质组分析。

2.2.2 数据非依赖性采集(DIA)技术

DIA技术的原理是首先将质谱扫描的整个一级质量范围根据离子密度分割为若干窗口,每个窗口依次选择、碎裂、扫描,DIA能够获得每个窗口范围内所有母离子的全部碎片离子信息,循环时间固定,同时数据可以回溯,有效解决了DDA技术存在的随机性和低丰度离子缺失的问题。但是DIA模式产生的碎片离子谱图过于复杂,丢失了母离子与碎片离子的对应关系,因此高度复杂的混合二级谱图对肽段和蛋白质的正确解析提出了挑战。为了减少肽段共流出问题,降低DIA谱图的复杂度,一些具有革命性突破的DIA策略也相继被开发。如diaPASEF技术^[30]通过离子淌度提供额外一个维度的分离,使DIA数据采集时在不牺牲窗口循环速度的同时,降低谱图复杂度和提高离子利用率,带来4D-蛋白质组学在鉴定深度、检测周期、定量准确性等性能的全方位提升。Scanning-SWATH技术^[31]利用四级杆的连续扫描功能,通过设定更小的质量隔离窗口和超高的扫描速度,提供更好的谱图专属性和灵敏度,提高定性、定量的准确度。综上,DIA技术是一种高通量的蛋白质组学研究手段,可应用于COVID-19大样本高通量差异蛋白质组学研究和翻译后修饰位点分析。

2.2.3 靶向数据采集技术

靶向数据采集技术针对目标蛋白/肽段离子进行靶向监测和采集,主要采集方式包括基于三重四级杆质谱仪的多反应监测(multiple reaction monitoring,MRM)和基于高分辨质谱仪的平行反应监测(parallel reaction monitoring,PRM)。MRM技术选择目标肽段的特定母离子和子离子对进行靶向质谱分析,最大限度排除干扰离子的影响,显著提高了目标肽段的信噪比,常用于高通量靶向蛋白质定量。该技术具有灵敏度高、准确性好、特异性强的优点,被誉为质谱定量的“金标准”,特别适用于蛋白质标志物的高通量验证^[32]。PRM是MRM的一种衍生技术,是基于高分辨、高精度质谱进行靶向定量,PRM选择目标肽段的特定母离子碎裂,对母离子得到的所有子离子进行全扫描。PRM基于高分辨质谱,精度更高、抗背景干扰能力更强^[33-34]。在应用方面,MRM/PRM技术可替代传统的Western blot技术,能够实现在大规

模生物样本中进行高通量的目标蛋白质验证,提高实验效率。在COVID-19研究中,MRM和PRM技术常用于发现的潜在生物标志物和药物靶标的验证。

3 蛋白质组学在COVID-19研究中的应用

基于质谱的蛋白质组学技术进行COVID-19研究的工作流程如图3所示,通过对病原体和感染后的宿主细胞、组织、体液等进行蛋白质组学研究,可以快速全面的筛选到重要的目标蛋白,并结合生物信息学分析、蛋白质相互作用分析等,为揭示COVID-19相关的生物标志物、研究分子机制、识别潜在药物靶点提供了一种有效的策略。

3.1 面向精准诊断的生物标志物研究

生物标志物是指一种可客观检测和评价的指标性物质,作为正常生物学过程、病理过程或治疗干预药理学反应的指示因子^[35]。生物标志物的发现对于疾病的早期诊断、治疗监测、疾病分级以及预后评估等领域具有重要的价值。随着质谱技术的进步以及生物信息学与统计学算法的发展,运用蛋白质组学方法寻找和发现有价值的生物标志物是目前研究的一个热点。其中鼻咽拭子、唾液、血液和尿液等体液样本具有低侵入性、易于获取、易于采集和处理、成本低等优点,成为人类疾病标志物研究的最佳样品来源^[36]。

3.1.1 鼻咽拭子检测和病毒颗粒标志物

鼻咽部位是SARS-CoV-2发生感染和免疫反应的“第一战场”,因此,鼻咽拭子、唾液样本等常用于进行病毒颗粒的检测和诊断标志物研究。病毒颗粒是由病毒的结构蛋白(S蛋白、M蛋白、E蛋白、N蛋白)等以及遗传物质RNA组成。采用蛋白质组学技术对SARS-CoV-2进行研究,可以获得病毒颗粒的蛋白质组成及其修饰信息,从而发现特征性蛋白和特征多肽,用于病毒检测和疾病的早期诊断(表1)。

基于MALDI-TOF MS的蛋白质组学技术用于SARS-CoV-2快速检测显示了较好的准确度。Nachtigall等^[24]采用MALDI-TOF MS技术结合机器学习的方法对362份鼻咽拭子样本(211份SARS-CoV-2阳性样本和151份阴性样本)进行检测。研究发现,支持向量机模型(support vector machines,SVM)分析得到了最高的准确率(93.9%),阳性样本的误判率为7%,阴性样本的误判率为5%。结果表明,MALDI-TOF MS和机器

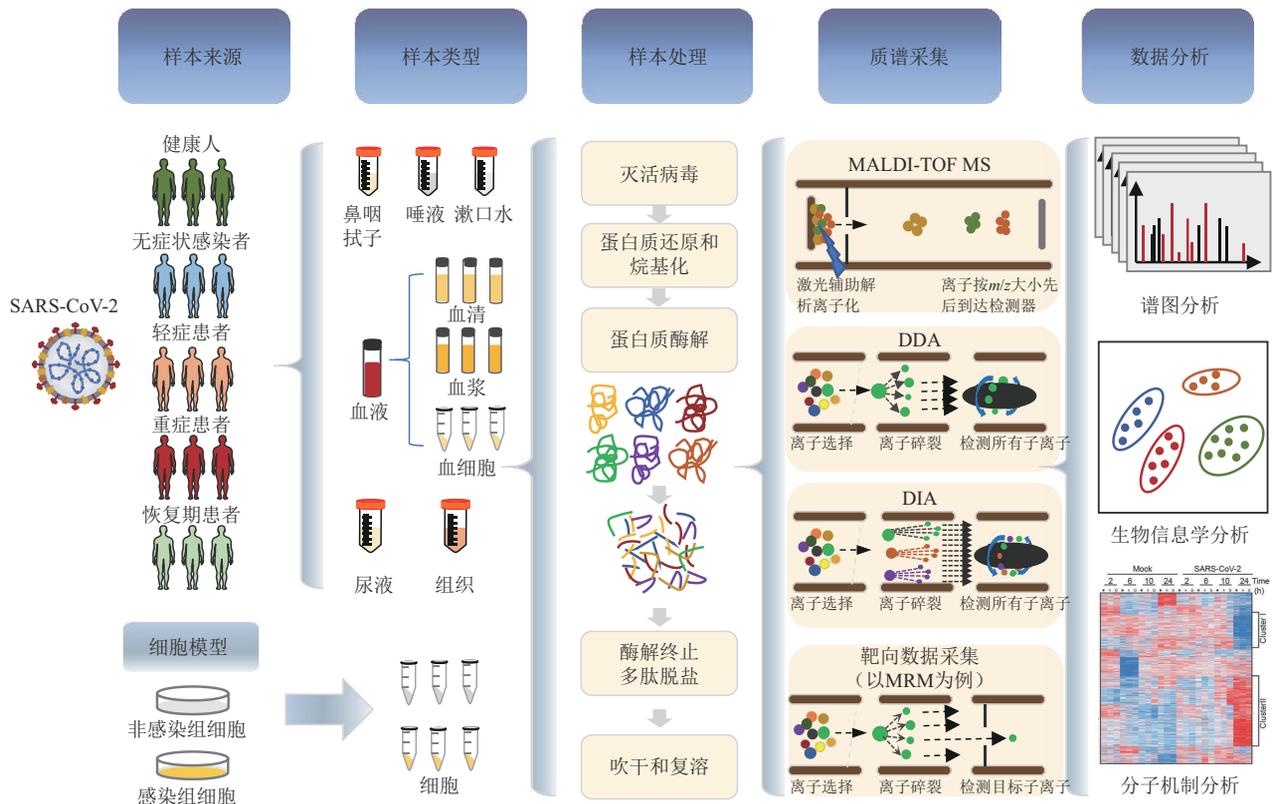


Fig. 3 Workflow based on recent MS-based proteomic analysis on COVID-19

图3 基于质谱的蛋白质组学技术进行COVID-19研究的工作流程

基于质谱的蛋白质组学技术进行COVID-19研究的工作流程: 样本来源主要为具有不同症状的患者或感染SARS-CoV-2的细胞模型; 采用的样本主要为来自感染患者或对照组的鼻咽拭子、唾液、漱口水、血液、尿液以及培养的细胞; 样本处理方式为蛋白质还原、烷基化和酶解得到多肽; 质谱采集方式主要通过MALDI-TOF MS技术和基于LC-ESI-MS的数据依赖性采集 (DDA)、数据非依赖性采集 (DIA) 和靶向数据采集技术; 最后根据采集的谱图比较后进行生物信息学分析和分子机制分析, 发现COVID-19相关的生物标志物、感染机制和治疗靶标。

学习分析可用于检测鼻拭子样本中的SARS-CoV-2。Iles等^[25]采用MALDI-TOF MS进行SARS-CoV-2病毒蛋白检测, 结果显示, 在假病毒、唾液和漱口水样本中均能检出S蛋白的S1和S2b片段, 其中S1片段的检测率和特异性接近100%。该方法取样容易、样品处理简单、分析速度快、准确率高, 这些特点揭示了MALDI-TOF MS技术的临床诊断应用价值。

高通量DDA/DIA技术结合PRM/MRM靶向定量的策略发现多个特征多肽可作为SARS-CoV-2病毒颗粒的检测标志物。Gouveia等^[37]采用DDA技术鉴定到101条来自SARS-CoV-2的特征肽段, 通过排除与其他物种有交叉和有变异的多肽, 最终选择S蛋白的3个特征肽 (FQTLLALHR、HTPINLVR、LQSLQTYVTQQLIR) 和M蛋白的1个特征肽 (VAGDSGFAAYS) 作为最佳特征肽段

可用于SARS-CoV-2感染的检测。Bezstarosti等^[38]通过DDA技术鉴定到9个SARS-CoV-2病毒蛋白; 通过PRM靶向定量技术在临床病人鼻咽拭子和痰液样本中定量到4个特征肽段: 3个来自N蛋白 (GFYAEGSR、ADETQALPQR、AYNVTQAFGR), 1个来自M蛋白的特征肽段可应用于SARS-CoV-2感染的检测。Cazares等^[39]建立了基于PRM技术检测SARS-CoV-2病毒S蛋白和N蛋白的检测方法, S蛋白的特征肽 (FQTLLALHR)、N蛋白的特征肽 (DQVILLNK) 可用于病毒检测。Cardozo等^[40]采用DDA技术结合PRM技术的策略建立高通量SARS-CoV-2临床质谱检测方法, 可直接从鼻咽和口咽拭子中检测SARS-CoV-2 N蛋白的肽段, 通过在985例临床样本中对该方法进行定性和定量验证, 准确度>87%, 特异性>95%。以上结果揭示了DDA/DIA技术结合PRM/MRM靶向定量的策略具

有高通量、高灵敏度、高准确度的特点，显示了MRM/PRM技术的临床诊断应用价值。

综合以上研究结果，目前用于鼻咽拭子检测的病毒颗粒标志物主要来源于SARS-CoV-2的结构蛋白N蛋白、M蛋白和S蛋白。有以下特征肽段可作为SARS-CoV-2感染诊断的标志序列：N蛋白的特征肽段GFYAEGSR、RGPEQTQGNFGDQELIR、

ADETQALPQR、DQVILLNK；M蛋白的特征肽段EITVATSR、VAGDSGFAAYSR；S蛋白的S1片段和特征肽段QIAPGQTGK、FQTLALHR、HTPINLVR、LQSLQTYVTQQLIR。这些多肽序列有望应用于SARS-CoV-2感染的临床筛查和病毒检测。

Table 1 Application of proteomics in the study of SARS-CoV-2 nasopharyngeal swab test and viral particle biomarkers

表1 蛋白质组学在SARS-CoV-2鼻咽拭子检测和病毒颗粒标志物研究中的应用

实验样本	技术方法	研究结果	参考文献
362例患者鼻拭子样本（211例阳性样本，151例阴性样本）	MALDI-TOF MS 技术	共检测到31个差异峰，SVM模型分析得到了最高的诊断准确率（93.9%），阳性样本的误判率为7%，阴性样本的误判率为5%	[24]
病毒模拟样本、唾液和漱口口水样本	MALDI-TOF MS 技术	在模拟样本、唾液和漱口口水样本中均能检出S蛋白的S1和S2b肽段，其中S1片段的检测率和特异性接近100%	[25]
SARS-CoV-2感染Vero细胞的模拟样本	DDA技术	鉴定到101个来自SARS-CoV-2的特征肽段，其中S蛋白的3个特征肽（FQTLALHR、HTPINLVR、LQSLQTYVTQQLIR），以及M蛋白的1个特征肽（VAGDSGFAAYSR）作为最佳特征肽段可用于SARS-CoV-2感染的检测	[37]
SARS-CoV-2侵袭Vero细胞的模拟样本，病人鼻咽拭子和痰液样本	DDA技术和PRM技术	通过在临床病人鼻咽拭子和痰液样本中进行检测，定量到4个特征肽段，3个来自N蛋白（GFYAEGSR、ADETQALPQR、AYNVTQAFGR），1个来自M蛋白（EITVATSR）的特征肽段可应用于检测COVID-19患者临床标本中的病毒蛋白	[38]
重组S蛋白和N蛋白，人肺上皮细胞黏液样本	DDA技术和PRM技术	建立了基于PRM技术的SARS-CoV-2病毒S蛋白和N蛋白检测方法，最后筛选S蛋白的特征肽（FQTLALHR），N蛋白的特征肽（DQVILLNK）用于病毒检测，检出限（limit of detection, LOD）达到195 amol，定量限（limit of quantitation, LOQ）达到390 amol	[39]
重组N蛋白，985例患者鼻咽和口咽拭子样本	DDA技术和PRM技术	采用PRM技术建立高通量SARS-CoV-2临床质谱检测方法，可直接从鼻咽和口咽拭子中检测SARS-CoV-2 N蛋白的肽段，实现500个样本/天的高通量检测。通过在985例临床样本中对该方法进行定性和定量验证，准确度>87%，特异性>95%	[40]

3.1.2 血浆/血清生物标志物

血液是临床检测中最常用的样本。血液由血浆和血细胞组成，血浆是血液中加入抗凝剂分离后的上清液；血清为未加入抗凝剂，经血液凝固后析出或离心后上层的液体，血清与血浆的主要区别是血清中不含纤维蛋白原，血浆和血清蛋白质表达的变化都可反映人体生理或病理状态，可用于疾病诊断、早期筛查以及预后评估等领域。通过对比COVID-19患者和健康人的血清/血浆蛋白质含量差别确定差异表达蛋白质，并分析相关的差异蛋白参与的分子通路和生物学功能（表2）。这些差异蛋白可作为诊断疾病和监测疾病进程的标志物，可以预测疾病的易感性和临床结局。

基于TMT标记定量的蛋白质组学研究发现多个潜在的不同疾病进程的COVID-19血浆/血清生物标志物。Shen等^[41]采用TMT标记定量技术进

行COVID-19患者血清的蛋白质组学和代谢组学表征，利用机器学习模型筛选到22个重要的蛋白质和7个代谢物，可用于进行严重COVID-19患者的预测和分类。这些差异化合物介导的病理生理途径包括免疫或炎症反应、血小板脱颗粒、凝血和代谢途径，其中最显著上调的是急性期蛋白，包括SAA1、SAA2、SAA4、CRP、SERPINA3、SAP/APCS等。Yan等^[42]结合蛋白质组学、代谢组学分析技术和患者临床生化检测指标发现，血清乳酸脱氢酶（LDH）水平与COVID-19的严重程度有关，血清LDH水平升高可能是缺氧和炎症引起的组织损伤的结果。Shu等^[43]采用TMT标记定量技术进行COVID-19血浆生物标志物研究，通过机器学习方法筛选了11个宿主蛋白可组合作为不同疾病进程的生物标志物，其中ORM1/AGP1、ORM2、FETUB和CETP可用于鉴别COVID-19患者和健康

对照; CETP、S100A9和CRP可用于预测COVID-19重症患者是否有致命的结局; AZGP1、ORM2和CFI可用于预测轻症到重症的临床结局; SERPINA3/ACT、LCPI/LPL和PI16的生物标志物组合可用于预测康复。

基于DIA技术的蛋白质组学研究发现, 多个潜在的血浆/血清生物标志物, 可用于进行严重COVID-19患者的预测和分类。Messner等^[44]基于DIA技术筛选到27种差异表达的潜在生物标志物可用于进行严重COVID-19患者的预测和分类, 其中补体系统蛋白(C1R、C1S、C8A、CFB、CFI、CFH)、急性反应期蛋白(CRP、SAA1、SAA2、HP)、白介素-6信号通路因子(ITIH4、LRG1、LBP)随疾病程度增加而上调, 而载脂蛋白(APOA1、APOC1)、ALB、GSN、TF随疾病程度增加而下调。Geyer等^[45]应用基于diaPASEF的4D-蛋白质组学技术检测COVID-19患者血清的54 d纵向时间表达变化, 结果显示先天免疫系统的蛋白质如CRP、SAA1、CD14、LBP和LGALS3BP在感染过程的早期下降, 25 d后又逐渐升高; 凝血调节剂(APOH、FN1、HRG、KNG1、PLG)和脂质稳态相关蛋白(APOA1、APOC1、APOC2、

APOC3、PON1)在疾病过程中逐渐升高。其中促炎急性期蛋白ITIH4在死亡病人取样第1天就显著升高, 其可能会成为预测死亡风险的生物标志物。Vollmy等^[46]采用基于DIA技术的蛋白质组学方法筛选到9个蛋白质生物标志物组合(HRG、FETUB、ITIH1、ITIH2、HPR、SERPINA3、LCAT、IGFALS、IGFBP3)用于死亡风险评估和预测。Chen等^[47]结合转录组学、蛋白质组学和代谢组学进行血浆生物标志物分析, 发现HGFAC、F13A1和CLEC3B可作为血浆蛋白质生物标志物。Messner等^[31]应用Scanning SWATH采集技术进行COVID-19血浆高通量蛋白质组学分析, 仅1 min梯度即可定量190个血浆蛋白质, 该研究新发现11个在COVID-19患者中差异表达的生物标志物, 主要为急性期反应和补体级联反应相关蛋白。

综合以上血浆/血清蛋白质组学研究结果, COVID-19血浆/血清生物标志物主要是与急性期反应、补体级联反应、炎症反应、免疫反应、凝血级联反应和脂质代谢相关的蛋白质, 其中急性期反应蛋白SAA1、CRP在多个研究中发现对预测疾病严重性进程具有重要意义。

Table 2 Application of proteomics in the study of COVID-19 plasma/serum biomarkers

表2 蛋白质组学在COVID-19血浆/血清生物标志物研究中的应用

实验样本	技术方法	研究结果	参考文献
46例COVID-19患者和53例对照个体的血清样本	TMT标记定量和PRM靶向采集技术, 代谢组学	采用机器学习模型筛选到22个重要的蛋白质和7个代谢物, 可用于进行严重COVID-19患者的预测和分类 差异化合物介导的病理生理途径, 包括免疫或炎症反应、血小板脱颗粒、凝血和代谢途径, 其中最显著上调的是急性期蛋白, 包括SAA1、SAA2、SAA4、CRP、SERPINA3、SAP/APCS等	[41]
144例COVID-19患者血清样本	TMT标记定量, 代谢组学, 临床生化检测	结合蛋白质组学、代谢组学技术和患者临床生化检测指标发现, 血清乳酸脱氢酶(LDH)水平与COVID-19的严重程度有关, 血清LDH水平升高可能是缺氧和炎症引起的组织损伤的结果	[42]
216例血浆样本(53例健康对照、56例轻症、53例重症、54例死亡)	TMT标记定量, ELISA验证	通过机器学习发现11个宿主蛋白可组合作为不同疾病进程的生物标志物, 其中ORM1/AGP1、ORM2、FETUB和CETP可用于COVID-19患者和健康志愿者的分类; CETP、S100A9和CRP可用于预测COVID-19重症患者是否有致命的结局; AZGP1、ORM2和CFI可用于预测轻症到重症的临床结局; SERPINA3/ACT、LCPI/LPL和PI16的生物标志物组合可用于预测康复	[43]
31例COVID-19患者血清样本	DIA技术	筛选到27种差异表达的潜在生物标志物可用于进行严重COVID-19患者的预测和分类, 其中补体系统蛋白(C1R、C1S、C8A、CFB、CFI、CFH), 急性反应期蛋白(CRP、SAA1、SAA2、HP), 白介素-6信号通路因子(ITIH4、LRG1、LBP)随疾病程度增加而上调, 而载脂蛋白(APOA1、APOC1)、ALB、GSN, TF随疾病程度增加而下调	[44]

续表2

实验样本	技术方法	研究结果	参考文献
720例人血清样本 (262例阴性对照样本, 458例COVID-19患者样本)	diaPASEF采集技术	结果显示先天免疫相关的蛋白质CRP、SAA1、CD14、LBP和LGALS3BP在感染过程的早期下降, 25 d后又逐渐升高; 凝血调节剂 (APOH、FN1、HRG、KNG1、PLG) 和脂质稳态相关蛋白 (APOA1、APOC1、APOC2、APOC3、PON1) 在疾病过程中逐渐升高 促炎急性期蛋白ITIH4在死亡病人取样第1天就显著升高, 其可能会成为预测死亡风险的生物标志物	[45]
33例COVID-19感染后 (17例幸存者和16例非幸存者) 病人的81个血清样本	DIA技术	筛选到9个蛋白质生物标志物组合 (HRG、FETUB、ITIH1、ITIH2、HPR、SERPINA3、LCAT、IGFALS、IGFBP3) 用于死亡风险评估和预测	[46]
66例COVID-19患者, 17例健康对照血液样本	DIA技术, 转录组学, 代谢组学	通过机器学习模型分析发现HGFAC、F13A1和CLEC3B可作为血浆蛋白质生物标志物	[47]
30例COVID-19患者, 15例健康对照血浆样本	Scanning SWATH技术	该研究建立了超高通量COVID-19血浆蛋白质组学分析方法, 仅1 min梯度即可定量190个血浆蛋白质, 研究新发现11个在COVID-19患者中差异表达的生物标志物, 包括急性期反应和补体级联反应等相关蛋白A2M、SERPINF2、PROS1、C1QC、IGKV4-1、SERPINA7、TTR、PON1、HPX、IGHG2、TMEM198	[31]

SAA1: 血清淀粉样蛋白A-1 (serum amyloid A-1 protein); SAA2: 血清淀粉样蛋白A-2 (serum amyloid A-2 protein); SAA4: 血清淀粉样蛋白A-4 (serum amyloid A-4 protein); CRP: C反应蛋白 (C-reactive protein); SERPINA3: α 1-抗胰凝乳蛋白酶 (α 1-antichymotrypsin); SAP/APCS: 血清淀粉样蛋白P成分 (serum amyloid P-component); ELISA: 酶联免疫吸附测定 (enzyme linked immunosorbent assay); ORM1/AGP1: α 1-酸性糖蛋白1 (α 1-acid glycoprotein 1); ORM2: α 1-酸性糖蛋白2 (α 1-acid glycoprotein 2); FETUB: 胎球蛋白B (fetuin-B); CETP: 胆固醇酯转运蛋白 (cholesterol ester transfer protein); S100A9: 钙结合蛋白S100A9 (protein S100-A9); AZGP1: 锌- α 2糖蛋白 (zinc- α 2-glycoprotein); CFI: 补体因子I (complement factor I); LCP1/LPL: 淋巴细胞胞浆蛋白1/L-丝束蛋白 (lymphocyte cytosolic protein 1/L-plastin); PI16: 肽酶抑制剂16 (peptidase inhibitor 16); LDH: 血清乳酸脱氢酶 (serum lactate dehydrogenase); C1R: 补体成分C1r (complement C1r); C1S: 补体成分C1s (complement C1s); C8A: 补体成分C8 (complement C8 alpha chain); CFB: 补体因子B (complement factor B); CFH: 补体因子H (complement factor H); HP: 触珠蛋白 (haptoglobin); ITIH4: 胰蛋白酶抑制剂重链H4 (inter-alpha-trypsin inhibitor heavy chain 4); LRG1: 富亮氨酸 α 2糖蛋白 (leucine-rich alpha-2-glycoprotein); LBP: 脂多糖结合蛋白 (lipopolysaccharide-binding protein); APOA1: 载脂蛋白A1 (apolipoprotein A-I); APOC1: 载脂蛋白C1 (apolipoprotein C-I); ALB: 白蛋白 (albumin); GSN: 凝溶胶蛋白 (gelsolin); TF: 转铁蛋白 (transferrin); CD14: 单核细胞分化抗原CD14 (monocyte differentiation antigen CD14); LGALS3BP: 半乳糖凝集素3结合蛋白 (galectin-3-binding protein); APOH: 载脂蛋白H/ β 2糖蛋白I (beta-2-glycoprotein 1); FN1: 纤连蛋白 (fibronectin); HRG: 富组氨酸糖蛋白 (histidine-rich glycoprotein); KNG1: 激肽原 (kininogen-1); PLG: 纤溶酶原 (plasminogen); APOC2: 载脂蛋白C2 (apolipoprotein C-II); APOC3: 载脂蛋白C3 (apolipoprotein C-III); PON1: 血清对氧磷酶/芳香酯酶1 (serum paraoxonase/arylesterase 1); ITIH1: 胰蛋白酶抑制剂重链H1 (inter-alpha-trypsin inhibitor heavy chain H1); ITIH2: 胰蛋白酶抑制剂重链H2 (inter-alpha-trypsin inhibitor heavy chain H2); HPR: 结合珠蛋白相关蛋白质 (haptoglobin-related protein); LCAT: 卵磷脂胆固醇脂酰转移酶 (phosphatidylcholine-sterol acyltransferase); IGFALS: 胰岛素样生长因子酸不稳定亚基 (insulin-like growth factor-binding protein complex acid labile subunit); IGFBP3: 胰岛素样生长因子结合蛋白3 (insulin-like growth factor-binding protein 3); HGFAC: 肝细胞生长因子激活剂 (hepatocyte growth factor activator); F13A1: 凝血因子XIII A链 (coagulation factor XIII A chain); CLEC3B: 四连接素蛋白 (tetranectin); A2M: 巨球蛋白 (alpha-2-macroglobulin); SERPINF2: 抗纤溶酶 (α 2-antiplasmin); PROS1: 维生素K依赖蛋白S (vitamin K-dependent protein S); C1QC: 补体成分C1q子组分C (complement C1q subcomponent subunit C); IGKV4-1: 免疫球蛋白 κ 链4-1 (immunoglobulin kappa variable 4-1); SERPINA7: 甲状腺素结合球蛋白 (thyroxine-binding globulin); TTR: 甲状腺素运载蛋白 (transthyretin); HPX: 血液结合素 (hemopexin); IGHG2: 免疫球蛋白 γ 2 (immunoglobulin heavy constant gamma 2); TMEM198: 跨膜蛋白198 (transmembrane protein 198)。

3.1.3 尿液生物标志物

尿液样本因非侵入性、易获取、可反映疾病的动态变化等优点, 在生物标志物研究中越来越受关注。尿液没有稳态调控机制, 能更早的反应疾病变

化^[48]。通过对比COVID-19患者和健康人的尿液蛋白质含量差别确定差异表达蛋白质, 并分析其参与的分子通路和相关生物学功能 (表3), 可以为早期诊断和疾病进程研究提供重要意义。

Table 3 Application of proteomics in the study of COVID-19 urine biomarkers**表3 蛋白质组学在COVID-19尿液生物标志物研究中的应用**

实验样本	技术方法	研究结果	参考文献
106例尿液样本 (27例健康对照、15例非COVID-19样本、44例非严重患者样本、20例严重患者样本)	TMT标记定量技术, PRM靶向采集技术, ELISA验证	结果显示尿液蛋白质组比血清蛋白质组有更高的检测灵敏度, 建立了基于20个尿液蛋白质的机器学习模型, 包括VPS36、CEL、FREM2、TNR、PTGFRN、MLEC、ADGRL1、OSTN、ICOSLG、NPR3、MELTF、PRSS2、SERPINI1、MADCAM1、CDH22、CDH19、TRHDE、SPOCK1、CD84、BTNL3, 这20个尿液蛋白质组合具有对COVID-19轻重型进行分类预测的潜力	[49]
37例尿液样本 (14例COVID-19患者样本, 13例非COVID-19病人样本, 10例健康对照样本)	diaPASEF技术	研究发现在COVID-19感染早期发生免疫抑制和紧密连接障碍, 其中MT1G、LPL、 β 2M、PRKACA、FOLR2和APOA4发生显著改变, 有望作为尿液生物标志物	[50]
40例尿液样本 (32例健康对照, 6例COVID-19患者样本, 2例恢复期样本)	DDA技术	研究发现COVID-19患者补体系统、缺氧反应分子上调; 而血小板脱颗粒、糖代谢和脂代谢下调 其中, 上调的蛋白HYOU1、D-dimer、SERPIND1和下调的蛋白NPC2、APOA1、CUBN有望作为尿液生物标志物	[51]
50例尿液样本 (39例COVID-19患者样本, 11例非COVID-19对照样本)	TMT标记定量技术	研究发现在COVID-19阳性样本中, 急性期反应、先天免疫反应、补体激活的调节和血小板激活过程相关蛋白上调; 细胞黏附、细胞外基质组织和白细胞迁移过程相关蛋白下调	[52]
19例严重患者尿液样本, 16例非严重患者样本	DIA技术, 代谢组学, 脂质组学	该研究基于多组学方法, 全面分析了COVID-19患者血浆和尿液中的分子变化, 发现了一系列生物标志物 在验证组尿液样本中, 生物标志物组合预测AUC值达到0.904, 尿液的精确预测证明了患者动态监测的可行性	[53]
1 925名成人的尿液样本, 284名健康儿童尿液样本	DDA/DIA技术	在1 925名成人的尿液样本中, 80.1%检测到ACE2并进行了定量; 284名健康儿童中14.1%的样本检测到ACE2。研究证明尿液中ACE2可以通过基于质谱的蛋白质组学方法检测, 尿液ACE2水平在大队列长期监测结果较稳定, 并且与多种生理和病理条件有关 ACE2有望作为预测SARS-CoV-2感染风险及其并发症的生物标志物	[54]

VPS36: 液泡蛋白-分选相关蛋白36 (vacuolar protein-sorting-associated protein 36); CEL: 胆盐激活脂肪酶抗体 (bile salt-activated lipase); FREM2: FRAS1相关的细胞外基质蛋白2 (FRAS1-related extracellular matrix protein 2); TNR: 肌腱蛋白 (tenascin-R); PTGFRN: 前列腺素F2受体负调节因子 (prostaglandin F2 receptor negative regulator); MLEC: Malectin 蛋白 (malectin); ADGRL1: 黏附G蛋白偶联受体L1 (adhesion G protein-coupled receptor L1); OSTN: 骨织素 (osteocrin); ICOSLG: 可诱导共刺激分子 (ICOS ligand); NPR3: 心房利钠肽受体3 (atrial natriuretic peptide receptor 3); MELTF: 黑素转铁蛋白 (melanotransferrin); PRSS2: 胰蛋白酶原2 (trypsin-2); SERPINI1: 神经源性丝氨酸蛋白酶抑制剂 (neuroserpin); MADCAM1: 黏膜地址素细胞黏附分子 (mucosal addressin cell adhesion molecule 1); CDH22: 钙黏蛋白22 (cadherin-22); CDH19: 钙黏蛋白19 (cadherin-19); TRHDE: 促甲状腺激素释放激素降解酶 (thyrotropin-releasing hormone-degrading ectoenzyme); SPOCK1: 睾丸蛋白聚糖1 (testican-1); CD84: SLAM家庭成员5 (SLAM family member 5); BTNL3: 嗜乳脂蛋白样蛋白3 (butyrophilin-like protein 3); MT1G: 金属硫蛋白-1G (metallothionein-1G); LPL: 脂蛋白脂肪酶 (lipoprotein lipase); β 2M: β 2微球蛋白 (beta-2-microglobulin); PRKACA: cAMP依赖的蛋白激酶催化亚基 (cAMP-dependent protein kinase catalytic subunit alpha); FOLR2: 叶酸受体 β (folate receptor beta); APOA4: 载脂蛋白A4 (apolipoprotein A-IV); HYOU1: 低氧上调蛋白1 (hypoxia up-regulated protein 1); D-dimer: D-二聚体; SERPIND1: 肝素辅因子2 (heparin cofactor 2); NPC2: NPC细胞内胆固醇转运蛋白2 (NPC intracellular cholesterol transporter 2); CUBN: 内皮维生素B12受体 (cubilin)。

Bi等^[49]全面分析了COVID-19患者血清和尿液的蛋白质组学和代谢组学数据, 结果显示尿液蛋白质组比血清蛋白质组有更高的检测灵敏度, 并建立了基于20个尿液蛋白质的生物标志物组合, 具有对COVID-19轻重型进行分类预测的潜力。Tian等^[50]采用基于diaPASEF的4D-蛋白质组学技术进

行COVID-19尿液生物标志物和感染机制研究, 发现在COVID-19感染早期发生免疫抑制和紧密连接障碍, 其中MT1G、LPL、 β 2M、PRKACA、FOLR2和APOA4发生显著改变, 有望作为尿液生物标志物。Li等^[51]采用DDA技术进行COVID-19尿液蛋白质组学分析, 发现COVID-19患者补体系

统、缺氧反应分子上调；而血小板脱颗粒、糖代谢和脂代谢下调。其中，上调的蛋白HYOU1、D-dimer、SERPIND1和下调的蛋白NPC2、APOA1、CUBN有望作为尿液生物标志物。Chavan等^[52]采用TMT标记定量技术进行尿液蛋白质组学分析，发现在COVID-19阳性样本中，急性期反应、先天免疫反应、补体激活的调节和血小板激活过程相关蛋白上调；细胞黏附、细胞外基质组织和白细胞迁移过程相关蛋白下调。Li等^[53]采用定量蛋白质组学和代谢组学方法全面分析了COVID-19患者血浆和尿液中的分子变化，发现了一系列生物标志物。在验证组尿液样本中，生物标志物组合预测AUC值达到0.904。Ni等^[54]分析了大量基于LC-MS/MS的尿液蛋白质组学数据，发现在1925名成人的尿液样本中80.1%的样本检测到ACE2，284名健康儿童中14.1%的样本检测到ACE2。研究证明，尿液中ACE2可以通过基于质谱的蛋白质组学方法检测，尿液ACE2水平在大队列长期监测结果中较稳定，并且与多种生理和病理条件有关。ACE2有望作为预测SARS-CoV-2感染风险及其并发症的生物标志物。

综合以上尿液蛋白质组学结果，尿液蛋白质组比血清蛋白质组有更高的检测灵敏度，COVID-19尿液生物标志物与血浆/血清生物标志物反映了相同的生物学通路的变化，包括急性期反应、补体级联反应、炎症反应、免疫系统、凝血和代谢途径等。尿液样本具有完全无创、可连续收集、更容易检测低丰度蛋白、早期发现、检测敏感、特异性强等优势^[49]，被认为是鉴定生物标志物的理想来源，并且在大队列长期监测结果中较稳定，有望作为COVID-19早期诊断和疾病进程研究的重要生物标志物来源^[55]。

3.2 面向SARS-CoV-2感染的分子机制研究

蛋白质组表达的改变不仅可以用于发现生物标志物，而且可以揭示疾病的发生发展变化过程。尽管现在对COVID-19进行了快速和广泛的研究，但对其发病机制，特别是病理生理条件下的分子机制仍然知之甚少。通过蛋白质组学技术可以快速全面的筛选到重要的目标蛋白，大大缩短实验周期，并结合生物信息学通路分析、蛋白质相互作用分析，为揭示与SARS-CoV-2致病性相关的分子机制提供了一种有效的策略。目前，基于蛋白质组学技术对SARS-CoV-2感染的分子机制研究主要采用病毒感染的细胞模型和患者组织或体液样本进行研究。

3.2.1 基于病毒感染的细胞模型研究

病毒感染可以通过操纵宿主代谢来启动，适应宿主代谢环境是病毒病原体复制、繁殖和战胜它们入侵的邻近细胞的先决条件^[56]。为了解感染后宿主蛋白质组的变化、深入认识发病机理、探究生物学机制，通常必须借助病毒感染细胞的模拟样本。Bojkova等^[57]通过研究病毒感染的细胞模型发现SARS-CoV-2感染后重塑了蛋白质翻译、剪接、碳水化合物、核酸代谢和蛋白质稳态等中心细胞途径。

蛋白质的磷酸化对蛋白质的功能具有重要影响，可以改变细胞的信号传导、发育和分化以及周期控制和代谢。Bouhaddou等^[58]采用Vero E6细胞模型进行了SARS-CoV-2感染后的磷酸化蛋白质组学研究，发现SARS-CoV-2感染调节宿主激酶信号，其中p38丝裂原活化蛋白激酶(p38/MAPK)、酪蛋白激酶II(CK2)、钙/钙调素依赖性蛋白激酶II(CAMK2G)等激酶活性上调，而周期蛋白依赖性激酶(CDK)、蛋白激酶B(AKT)以及Rho家族的激酶活性出现下调。结合细胞学和生物化学的实验证明，病毒N蛋白能够同CK2结合并进而调节细胞骨架的组装，而病毒侵染会导致p38/MAPK的升高并导致细胞周期中止。Stukalov等^[59]用SARS-CoV-2、SARS-CoV分别感染A549肺癌细胞，分析了病毒对宿主细胞转录组、蛋白质组、泛素化和磷酸化修饰组的影响。该研究共定量到16399个磷酸化位点，其中4643个磷酸化位点在病毒感染后出现显著变化，与细胞存活、细胞周期进展、细胞生长和动力、应激反应和DNA损伤反应有关的中心激酶(CDKs、AKT、MAPKs、ATM和CHEK1)参与了调控。细胞相互作用分析结果显示，SARS-CoV-2调节多种细胞功能，如先天免疫调节通路、胆固醇代谢等，SARS-CoV和SARS-CoV-2病毒感染均导致I型干扰素反应下调并激活促炎信号。

蛋白质糖基化修饰作为最常见也是最重要的一种翻译后修饰种类，在病毒的生命周期中起着不可替代的作用，包括调节蛋白质折叠、干预受体结合、影响蛋白质降解率、调控宿主偏好和屏蔽免疫系统识别的免疫原性表位等^[60-61]。Watanabe等^[60]研究表达于293F细胞系SARS-CoV-2的S蛋白，发现22个N-糖基化位点均发生N-糖基化修饰，糖链类型包括高甘露糖型、杂合型和复杂型，其中高甘露糖型和复杂型糖链占比较高，这些位点上15%的糖链含至少一个唾液酸残基，52%的糖链发生了

岩藻糖基化。Dong 等^[62]采用 LC-MS/MS 技术对 SARS-CoV-2 的 S 蛋白进行 O-糖基化修饰分析, 鉴定到 27 个 O-糖基化位点和 66 种 O-糖链结构, 其中 26 种含唾液酸残基、37 种发生了岩藻糖基化。Tian 等^[61]采用基于质谱的糖蛋白质组学技术, 从 SARS-CoV-2 病毒颗粒中提取的 S 蛋白上鉴定到了 17 个 O-糖基化修饰位点, 其中有 11 个位点发生在 N-糖基化位点 1 至 3 个氨基酸附近 (“N±1~3”), 其中的 4 个位点发生了唾液酸修饰 (T236、S659、T1076 和 T1077), 通过定点突变验证 N-糖基化位点是这些 O-糖基化位点出现的先决条件, 并将这一现象命名为 “O-Follow-N” 规律 (表 4)。糖基化修饰对于 COVID-19 发病机制的探索、疫苗和治疗药物的设计开发, 以及检测试剂盒的生产具有重要意义。

Wang 等^[63]采用 4D 蛋白质组学技术和生物成像技术, 通过模拟肺泡-微血管屏障, 揭示了 SARS-CoV-2 感染对于肺泡上皮细胞和肺微血管内皮细胞的损伤机理。在肺泡上皮细胞中, 细胞周期、细胞增殖和细胞程序性死亡等生物学过程受到显著影响。SARS-CoV-2 感染肺泡上皮细胞后, 可激活其抗病毒和先天免疫反应, 上调促炎细胞因子表达, 如 IL-1 α 和干扰素。这些细胞因子释放到细胞外空间, 并进一步诱发相邻肺微血管内皮细胞的损伤, 最终导致肺泡-微血管屏障破坏。

综合以上基于病毒感染的细胞模型研究结果, 揭示了 SARS-CoV-2 感染细胞后重塑了蛋白质翻译、剪接、碳代谢、核酸代谢和蛋白质稳态等中心细胞途径, 细胞周期、细胞增殖和细胞程序性死亡等生物学过程受到显著影响。

Table 4 Partial reports of N-glycosylation and O-glycosylation of SARS-CoV-2 spike protein based on glycoproteomics

表4 基于糖蛋白质组学技术的SARS-CoV-2刺突蛋白N-糖基化和O-糖基化修饰的部分报道

糖基化类型	位点数	具体位点	研究结果	参考文献
N-糖基化修饰	22	N17、N61、N74、N122、N149、N165、N234、N282、N331、N343、N603、N616、N657、N709、N717、N801、N1074、N1098、N1134、N1158、N1173、N1194	刺突蛋白22个N-糖基化位点均发生N-糖基化修饰, 共鉴定到110种糖链结构。其中2个位点N234和N709主要为高甘露糖, 在N61、N122、N603、N717、N801和N1074位点可以鉴定到寡甘露糖型和复合型聚糖的混合物	[60]
O-糖基化修饰	26	T73、S221、T284、T323、S325、T547、T618、T630、T632、S673、T716、S803、S810、S813、S975、S1123、T1136、S1175	鉴定到27个O-糖基化位点, 其中18个O-糖基化位点被明确识别, 另外9个没有得到明确的O-糖基化位点。在S1和S2亚基上分别鉴定到15个和12个O-糖基化位点。共鉴定到66种O-糖链结构, 其中26种含唾液酸残基、37种发生了岩藻糖基化	[62]
O-糖基化修饰	17	T22、S60、T124、S151、T236、S305、T307、T323、T604/S605、T618、S659、T696、T724、T1076、T1077、S1097、T1100	定量结果显示SARS-CoV-2 S蛋白不同O-糖基化位点上丰度前5名的 O-连接聚糖依次为HexNAc (2) Hex (3)、HexNAc (2) Hex (2)、HexNAc (1) Hex (1) NeuAc (2)、HexNAc (1) Hex (1) 和 HexNAc (1)	[61]
O-糖基化修饰	17	T22、S60、T124、S151、T236、S305、T307、T323、T604/S605、T618、S659、T696、T724、T1076、T1077、S1097、T1100	鉴定到了17个O-糖基化修饰位点, 其中有11个位点发生在N-糖基化位点1至3个氨基酸附近 (“N±1~3”), 其中的4个位点发生了唾液酸修饰 (T236、S659、T1076和T1077), 通过定点突变验证N-糖基化位点是这些O-糖基化位点出现的先决条件, 并将这一现象命名为 “O-Follow-N” 规律	[61]

3.2.2 基于患者组织或体液样本研究

通过对 COVID-19 患者的组织和体液样本的蛋白质组学分析, 可以快速全面的得到所有蛋白质变

化的信息, 为揭示与 SARS-CoV-2 致病性相关的分子机制提供了一种有效的策略。目前研究结果显示, COVID-19 患者中与免疫反应和炎症反应、代

谢途径、凝血级联反应等通路相关的蛋白质严重失调。

COVID-19患者适应性免疫应答降低,炎症反应的多个通路被激活。Nie等^[20]通过蛋白质组学技术,系统分析了COVID-19去世患者7种器官的组织病理学改变。在肺中检测到两种免疫关键蛋白癌胚抗原相关细胞黏附分子1(CEACAM1)和CD276蛋白上调,在肺和脾中检测到淋巴细胞特异性酪氨酸蛋白激酶(LCK)下调,表明T细胞介导的免疫反应在肺和脾脏受到抑制;同时脾脏中T细胞衰竭和单核细胞生物标志物上调提示高炎症反应;在肾皮质中检测到涉及炎症反应的多个通路被激活,包括LPS/IL-1介导的RXR功能抑制、急性期反应、toll样受体信号、IL-6信号和NF- κ B信号。Li等^[64]采用定量蛋白质组学方法对COVID-19病人外周血单核细胞(PBMCs)进行分析,发现COVID-19患者体内的T细胞激活、中性粒细胞介导的免疫反应通路发生显著变化,通路分析结果揭示了病毒蛋白与细胞炎症和先天性免疫反应相关蛋白的特异性相互作用,其中Nsp10与NF- κ B抑制因子(NKRF)相互作用促进IL-8的诱导,这有助于IL-8介导的中性粒细胞趋化和COVID-19患者过度炎症反应。Chen等^[47]对不同疾病程度的COVID-19患者以及健康对照血液样本进行蛋白质组学分析。结果显示,轻症和重症患者免疫反应表现不同,SARS-CoV-2在轻症患者的特点是入院时T细胞信号激活和T细胞分化富集,随后迅速减少;在重症患者中持续检测的T细胞信号呈阴性。重症患者在整个住院期间,IFN信号持续被激活,而先天性免疫信号的负调控因子(如TRIM59、USP21和NLRC3)下调。与轻症患者相比,重症患者IL-6、IL-8和IL-10水平显著升高。

COVID-19患者体内代谢途径紊乱,脂蛋白代谢、糖酵解和三羧酸循环(TCA循环)途径失调,脂肪酸 β 氧化和氧化磷酸化被激活。Chen等^[47]通过血液多组学分析结果显示,轻度和重度COVID-19患者的可溶性低密度脂蛋白受体(sLDLR)、卵磷脂胆固醇酰基转移酶(LCAT)、胆固醇酯转移蛋白(CETP)等脂蛋白代谢相关关键蛋白水平明显降低;COVID-19患者TCA循环周期中的酶如ACO2、IDH、OGDH、DLD、SDH和MDH较低,而脂肪酸合成的关键酶(乙酰辅酶A羧化酶(ACAC)和脂肪酸合成酶(FASN))升高;另外,血浆乳酸和乳酸脱氢酶也同时显著升高。这些

数据揭示了SARS-CoV-2感染导致脂蛋白代谢、糖酵解和TCA循环的失调。Nie等^[20]对COVID-19去世患者7种器官的蛋白质组学研究发现多种代谢过程被抑制,包括糖原分解、半乳糖降解和糖酵解;而在大多数器官中,脂肪酸 β 氧化和氧化磷酸化被激活,这表明一种转换到高效率的能量生产模式。Shen等^[41]对COVID-19患者血浆蛋白质组学和代谢组学分析,发现多种载脂蛋白的失调,包括APOA1、APOA2、APOH、APOL1、APOD和APOM。载脂蛋白在胆固醇反向转运中起重要作用,揭示了胆固醇代谢失调。Chen等^[65]分析了COVID-19患者在发病和康复阶段的血清蛋白质组,结果显示COVID-19患者在疾病期和恢复期胆固醇代谢和心肌功能相关通路长期紊乱。

COVID-19患者发生凝血功能障碍。COVID-19患者血液临床凝血试验结果显示D-dimer、凝血酶原时间(PT)、活化部分凝血活酶时间(APTT)上调,揭示血小板脱颗粒和凝血级联功能障碍与COVID-19的严重程度密切相关^[43, 66-67]。D'Alessandro等^[68]对COVID-19患者的血清蛋白质组分析,结果显示凝血/纤溶级联反应相关蛋白,包括SERPINA1、SERPINA3、SERPINF2等丝氨酸蛋白酶抑制剂和羧肽酶(CPB2/TAFI)的蛋白表达水平显著增加。Nie等^[20]对COVID-19患者7种器官的蛋白质组学研究数据揭示了参与凝血、抗凝和纤溶系统的多种蛋白质失调,其中凝血酶原(F2)、凝血因子XI、XII、XIII(A、F11、F12和F13A1)在COVID-19患者中均出现异常;纤维蛋白原 α 链、 γ 链和 β 链增加(FGA、FGG和FGB),这些蛋白质可以分裂成纤维蛋白,从而形成血凝块;在COVID-19器官中检测到多种丝氨酸蛋白酶抑制剂的失调:肝素辅因子2(SERPIND1)是一种凝血酶的抑制剂和肝素的辅因子,纤溶酶原激活物抑制剂(SERPINE1)是一种主要的纤溶酶抑制剂,可以破坏血液凝块。肾皮质中SERPIND1的下调和SERPINE1的上调可能是导致COVID-19肾皮质微血栓形成的原因。同时,COVID-19患者肺、肾、下肢静脉均可见微血栓,微血栓的形成是凝血、抗凝和纤溶系统之间的不平衡造成的。

综合以上蛋白质组学研究结果,揭示了COVID-19感染患者中免疫反应和炎症反应、代谢紊乱、凝血级联反应相关的蛋白质严重失调,通过影响上述通路而引发机体功能障碍和代谢紊乱是SARS-CoV-2感染机制之一。

3.3 面向精准治疗的药物靶标研究

目前, COVID-19 的治疗药物作用机制主要可以归纳为两大类: 一类是阻止病毒和宿主细胞结合, 作用的靶位是 S 蛋白、ACE2、TMPRSS2 等, 如目前处于临床试验阶段的该类药物有 TMPRSS2 抑制剂甲磺酸卡莫他司、单克隆抗体药物等; 另一类是阻止新病毒在宿主细胞内的产生, 抑制病毒复制和翻译过程, 作用的靶位是 RdRp 或者 3CLpro 等, 如目前处于临床试验阶段的该类药物有瑞德西韦、法维拉韦、利巴韦林、洛匹那韦和法匹拉韦等^[15]。尽管这些在体外显示出抗病毒活性的药物已进入临床研究, 但是一些药物临床试验结果显示对重症患者无效或治疗价值有限。到目前为止, 还没有治疗 COVID-19 的特效药物^[69]。蛋白质组学研究可以辅助绘制病原体在感染、入侵、持久性和发病机制后对宿主引起的变化的全景图, 并有助于确定新的治疗靶点^[53]。

Bojkova 等^[57] 研究证明, SARS-CoV-2 重塑了人类细胞中的翻译、剪接、碳代谢、核酸代谢和蛋白质稳态等中心细胞途径。该研究测试了多个关键途径的抑制剂。a. 翻译抑制剂: 环己亚胺 (翻译伸长抑制剂) 和依米汀 (40S 核糖体蛋白 S14 抑制剂) 可抑制 SARS-CoV-2 在细胞中的复制; b. 剪接体抑制剂: 以剪接因子 SF3B117 为靶点的剪接体抑制剂 pladienolide B, 可在对人类 Caco-2 细胞无毒的浓度下抑制 SARS-CoV-2 病毒复制; c. 糖酵解抑制剂: 己糖激酶的抑制剂 2-脱氧-D-葡萄糖可阻断糖酵解, 抑制 SARS-CoV-2 在 Caco-2 细胞中的复制; d. 核苷酸合成抑制剂: 鸟苷核苷酸从头合成中的限速酶肌苷单磷酸脱氢酶 (IMPDH) 的抑制剂利巴韦林, 可抑制 SARS-CoV-2 在细胞中的复制; e. 蛋白酶稳态扰动抑制剂: p97 是蛋白质稳态的关键成分, p97 的抑制剂 NMS-873 在低纳米摩尔浓度下抑制 SARS-CoV-2 复制。该研究为 COVID-19 的治疗提供了新的潜在靶点。Bouhaddou 等^[58] 通过分析受感染细胞的磷酸化图谱, 揭示现有的激酶靶向药物有望阻止 SARS-CoV-2 劫持宿主细胞。通过对全景磷酸化图谱中失调的激酶和通路分析, 找到 87 种相关的药物和化合物。通过测试发现, p38、CK2、CDKs、AXL 和 PIKFYVE 等激酶家族的抑制剂显示出了强大的抗病毒活性, 对 COVID-19 的治疗方法提供了新的思路 and 依据。

Gordon 等^[70] 鉴定了 332 个 SARS-CoV-2 与人类蛋白质之间的高可信度蛋白质-蛋白质相互作用,

这些相互作用与蛋白质贩运、翻译、转录和泛素化调控等多个生物学过程有关。以这些相互作用为靶点, 筛选到两类有效降低 SARS-CoV-2 病毒感染的分子: mRNA 翻译抑制剂 (zotatifin、ternatin-4 和 PS3061) 和 Sigma1、Sigma2 受体的调节剂 (氟哌啶醇、PB28、PD-144418 和羟氯喹)。研究发现, PB28 的效力是羟氯喹的 20 倍, mRNA 翻译抑制剂显示出较强的病毒抑制作用, 使得这类化合物成为极具吸引力的潜在药物。Stukalov 等^[59] 以病毒-宿主蛋白相互作用为靶点, 测试了 48 种药物对 SARS-CoV-2 的影响。其中用于治疗癌症和自身免疫性疾病的 B-RAF 抑制剂 (索拉非尼、瑞戈非尼、达拉非尼)、JAK1/2 抑制剂 (巴瑞替尼) 和 MAPK 抑制剂 (SB-239063) 可导致病毒生长的显著增加; 而 DNA 损伤诱导物 (替拉扎明、Rabusertib) 和 mTOR 抑制剂 (雷帕霉素) 可抑制病毒生长; FLT3 酪氨酸激酶抑制剂 (富马酸吉列替尼)、AKT 抑制剂 (Ipatasertib)、MMPs 抑制剂 (普马司他和马立马司他) 的抗病毒活性最高。这些化合物对 SARS-CoV-2 的复制具有明显的抑制作用, 但对细胞生长无影响或影响较小。这些研究基于宿主-病毒潜在的蛋白质相互作用分析, 为后续进行药物筛选和开发抗 SARS-CoV-2 相关药物研究提供了新的思路和策略。

明确药物靶标是治疗 COVID-19 的必由之路。21 世纪以来已经发生了 3 次由冠状病毒引起的重大传染病疫情, 但截止目前仍未发现特效治疗药物。SARS-CoV-2 未来可能会在人类社会长期存在, 成为全人类需要面对的共同挑战。因此, 需要开发针对 SARS-CoV-2 及广谱冠状病毒的特效药物。蛋白质组学在寻找药物靶标方面具有高通量、高灵敏的优势, 是未来筛选和开发抗 SARS-CoV-2 药物的重要技术手段。

4 总结与展望

自 COVID-19 大流行以来, 应用基于质谱的蛋白质组学技术进行 COVID-19 生物标志物研究、疾病机制研究、药物靶点筛选的研究与应用已经展现巨大的潜力。采用蛋白质组学技术在鼻咽拭子等样本中成功鉴定和定量来自 SARS-CoV-2 病毒颗粒的多个特征肽段, 这些特征肽段可作为潜在的 SARS-CoV-2 感染快速筛查和诊断的标志物; 高通量 DDA/DIA 技术结合 PRM 靶向定量的策略, 从 COVID-19 患者的血清/血浆或尿液样本中筛选与感

染相关的蛋白质,发现可作为诊断疾病和监测疾病进程的标志物,可以预测疾病的易感性和临床结局;蛋白质翻译后修饰研究、SARS-CoV-2与宿主细胞相互作用蛋白质组学研究,为揭示与SARS-CoV-2致病性相关的分子机制提供了一种有效的策略;基于蛋白质组学进行病毒-宿主蛋白质相互作用分析,在寻找药物靶标方面具有高通量、高灵敏的优势,为抗SARS-CoV-2药物研究提供重要的技术手段。

虽然应用蛋白质组学技术在COVID-19研究中取得较大进展,但仍有一些亟待解决的技术瓶颈,例如:a.目前质谱技术的自动化程度还有待提高,前处理过程也相对复杂,需要建立自动化和通量化的微量临床样本前处理技术体系;b.目前临床微量样本的质谱数据采集过程涉及较多步骤和参数,需要建立完整的行业标准以规范实验过程,保证实验间的重复性;c.针对大队列临床微量样本的蛋白质组学数据解析,还需要进一步提升鉴定和定量软件的灵敏度和准确度,发现低丰度蛋白质和标志物;d.蛋白质翻译后修饰(如S蛋白的糖基化以及蛋白质组的磷酸化、泛素化等)较复杂,样品中翻译后修饰蛋白质含量低、动态范围广给相关研究带来了很大的挑战,需要提高大规模翻译后修饰位点的鉴定和定量的能力;e.目前发现的蛋白质标志物和药物靶点均处于基础研究阶段,需要结合生物学和免疫学方法,实现候选蛋白分子的规模化验证。更先进的蛋白质组学技术必将促进COVID-19的精准诊疗研究。

随着质谱仪器性能和算法软件的不断发展与革新、相关技术瓶颈的不断突破,蛋白质组学数据采集和数据分析全流程标准化体系的建立,其应用的广度及深度将会迅速发展。蛋白质组学样品制备自动化技术逐渐成熟,可以实现短时间内对数以千计样本的快速前处理;随着大队列临床蛋白质组学的发展,生物标志物走向临床已成为发展趋势,单次分析可同时精确地检测出几十个甚至上百个生物标志物,并可检测出多种传统诊断技术无法检测到的生物标志物;将修饰蛋白质组学及蛋白质相互作用方法学应用于临床研究,实现修饰位点和药物靶点的精准检测,将拓展蛋白质分子机制的研究和药物靶点的发现。随着SARS-CoV-2突变株的发现,COVID-19发病机制愈加复杂,其精准诊疗面对更严峻的挑战,蛋白质组学技术有望走向临床成为广泛接受的COVID-19诊疗手段。

参 考 文 献

- [1] World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [EB/OL]. Geneva: World Health Organization [2020-03-11]. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- [2] Worldometers. Coronavirus Update (Live): 466,535,056 cases and 6,088,152 deaths from COVID-19 virus pandemic-worldometer [EB/OL]. New York: Worldometer [2022-03-18]. <https://www.worldometers.info/coronavirus/>
- [3] Wilkins M R, Sanchez J C, Gooley A A, *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*, 1996, **13**:19-50
- [4] Kahn P. From genome to proteome: looking at a cell's proteins. *Science*, 1995, **270**(5235):369-370
- [5] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*, 2020, **5**(4): 536-544
- [6] Lu R, Zhao X, Li J, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, 2020, **395**(10224): 565-574
- [7] Zhou P, Yang X L, Wang X G, *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020, **579**(7798): 270-273
- [8] Zhu N, Zhang D, Wang W, *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, 2020, **382**(8): 727-733
- [9] Wu A, Peng Y, Huang B, *et al.* Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*, 2020, **27**(3): 325-328
- [10] Wang Q, Zhang Y, Wu L, *et al.* Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*, 2020, **181**(4): 894-904
- [11] Hamming I, Timens W, Bulthuis M, *et al.* Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol*, 2004, **203**(2): 631-637
- [12] Hartenian E, Nandakumar D, Lari A, *et al.* The molecular virology of coronaviruses. *JBiol Chem*, 2020, **295**(37): 12910-12934
- [13] Kupferschmidt K, Cohen J. Race to find COVID-19 treatments accelerates. *Science*, 2020, **367**(6485): 1412-1413
- [14] V'kovski P, Kratzel A, Steiner S, *et al.* Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*, 2021, **19**(3): 155-170
- [15] Hu B, Guo H, Zhou P, *et al.* Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, 2021, **19**(3): 141-154
- [16] Liu F, Long X, Zhang B, *et al.* ACE2 expression in pancreas may cause pancreatic damage after SARS-CoV-2 infection. *Clin*

- Gastroenterol Hepatol, 2020, **18**(9): 2128-2130
- [17] Chen L, Li X, Chen M, *et al.* The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. *Cardiovasc Res*, 2020, **116**(6): 1097-1100
- [18] Zou X, Chen K, Zou J, *et al.* Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med*, 2020, **14**(2): 185-192
- [19] Li M Y, Li L, Zhang Y, *et al.* Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty*, 2020, **9**(1): 45
- [20] Nie X, Qian L, Sun R, *et al.* Multi-organ proteomic landscape of COVID-19 autopsies. *Cell*, 2021, **184**(3): 775-791
- [21] Varga Z, Flammer A J, Steiger P, *et al.* Endothelial cell infection and endotheliitis in COVID-19. *Lancet*, 2020, **395**(10234): 1417-1418
- [22] Sungnak W, Huang N, Becavin C, *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med*, 2020, **26**(5): 681-687
- [23] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, **422**(6928): 198-207
- [24] Nachtigall F M, Pereira A, Trofymchuk O S, *et al.* Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat Biotechnol*, 2020, **38**(10): 1168-1173
- [25] Iles R K, Zmuidinaite R, Iles J K, *et al.* Development of a clinical MALDI-TOF mass spectrometry assay for SARS-CoV-2: rational design and multi-disciplinary team work. *Diagnostics (Basel)*, 2020, **10**(10): 746
- [26] Fenn J B, Mann M, Meng C K, *et al.* Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 1989, **246**(4926): 64-71
- [27] Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*, 2016, **537**(7620): 347-355
- [28] 张伟, Kiyonami R, 江峥, 等. 创新数据非依赖性采集用于复杂基质目标蛋白质的定量分析. *分析化学*, 2014, **42**(12): 1750-1758
Zhang W, Kiyonami R, Jiang Z, *et al.* *Chinese J Anal Chem*, 2014, **42**(12): 1750-1758
- [29] 牟永莹, 顾培明, 马博, 等. 基于质谱的定量蛋白质组学技术发展现状. *生物技术通报*, 2017, **33**(9): 73-84
Mu Y Y, Gu P M, Ma B, *et al.* *Biotechnology Bulletin*, 2017, **33**(9): 73-84
- [30] Meier F, Brunner A D, Frank M, *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods*, 2020, **17**(12): 1229-1236
- [31] Messner C B, Demichev V, Bloomfield N, *et al.* Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol*, 2021, **39**(7): 846-854
- [32] Rifai N, Gillette M A, Carr S A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*, 2006, **24**(8): 971-983
- [33] Peterson A C, Russell J D, Bailey D J, *et al.* Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics*, 2012, **11**(11): 1475-1488
- [34] Schilling B, Maclean B, Held J M, *et al.* Multiplexed, scheduled, high-resolution parallel reaction monitoring on a full scan QqTOF instrument with integrated data-dependent and targeted mass spectrometric workflows. *Anal Chem*, 2015, **87**(20): 10222-10229
- [35] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 2001, **69**(3): 89-95
- [36] Veenstra T D, Conrads T P, Hood B L, *et al.* Biomarkers: mining the biofluid proteome. *Mol Cell Proteomics*, 2005, **4**(4): 409-418
- [37] Gouveia D, Grenga L, Gaillard J C, *et al.* Shortlisting SARS-CoV-2 peptides for targeted studies from experimental data-dependent acquisition tandem mass spectrometry data. *Proteomics*, 2020, **20**(14): e2000107
- [38] Bezstarosti K, Lamers M M, Doff W A S, *et al.* Targeted proteomics as a tool to detect SARS-CoV-2 proteins in clinical specimens. *PLoS One*, 2021, **16**(11): e0259165
- [39] Cazares L H, Chaerkady R, Samuel Weng S H, *et al.* Development of a parallel reaction monitoring mass spectrometry assay for the detection of SARS-CoV-2 spike glycoprotein and nucleoprotein. *Anal Chem*, 2020, **92**(20): 13813-13821
- [40] Cardozo K H M, Lebkuchen A, Okai G G, *et al.* Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nat Commun*, 2020, **11**(1): 6201
- [41] Shen B, Yi X, Sun Y, *et al.* Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell*, 2020, **182**(1): 59-72
- [42] Yan H, Liang X, Du J, *et al.* Proteomic and metabolomic investigation of serum lactate dehydrogenase elevation in COVID-19 patients. *Proteomics*, 2021, **21**(15): 2100002
- [43] Shu T, Ning W, Wu D, *et al.* Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity*, 2020, **53**(5): 1108-1122
- [44] Messner C B, Demichev V, Wendisch D, *et al.* Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst*, 2020, **11**(1): 11-24
- [45] Geyer P E, Arend F M, Doll S, *et al.* High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion. *EMBO Mol Med*, 2021, **13**(8): e14167
- [46] Vollmy F, Van Den Toorn H, Zenezini Chiozzi R, *et al.* A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Sci Alliance*, 2021, **4**(9): e202101099
- [47] Chen Y M, Zheng Y, Yu Y, *et al.* Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *EMBO J*, 2020, **39**(24): e105896
- [48] 井健, 高友鹤. 尿液作为新型生物标志物来源的探索. *生物化学与生物物理进展*, 2016, **43**(11): 1019-1028
Jing J, Gao Y H. *Prog Biochem Biophys*, 2016, **43**(11): 1019-1028

- [49] Bi X, Liu W, Ding X, *et al.* Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. *Cell Rep*, 2022, **38**(3):110271
- [50] Tian W, Zhang N, Jin R, *et al.* Immune suppression in the early stage of COVID-19 disease. *Nat Commun*, 2020, **11**(1): 5859
- [51] Li Y, Wang Y, Liu H, *et al.* Urine proteome of COVID-19 patients. *Urine (Amst)*, 2020, **2**:1-8
- [52] Chavan S, Mangalparthi K K, Singh S, *et al.* Mass spectrometric analysis of urine from COVID-19 patients for detection of SARS-CoV-2 viral antigen and to study host response. *J Proteome Res*, 2021, **20**(7): 3404-3413
- [53] Li Y, Hou G, Zhou H, *et al.* Multi-platform omics analysis reveals molecular signature for COVID-19 pathogenesis, prognosis and drug target discovery. *Signal Transduct Target Ther*, 2021, **6**(1): 155
- [54] Ni X, Sun C, Tian Y, *et al.* Could urinary ACE2 protein level help identify individuals susceptible to SARS-CoV-2 infection and complication?. *Sci China Life Sci*, 2020, **63**(11): 1766-1767
- [55] Wang S, Wong C C. A matter of perspective-cutting-edge technology-driven urine proteome in COVID-19. *Urine (Amst)*, 2021, **3**:1-2
- [56] Mahmud I, Garrett T J. Mass spectrometry techniques in emerging pathogens studies: COVID-19 perspectives. *J Am Soc Mass Spectrom*, 2020, **31**(10): 2013-2024
- [57] Bojkova D, Klann K, Koch B, *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*, 2020, **583**(7816): 469-472
- [58] Bouhaddou M, Memon D, Meyer B, *et al.* The global phosphorylation landscape of SARS-CoV-2 infection. *Cell*, 2020, **182**(3): 685-712
- [59] Stukalov A, Girault V, Grass V, *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature*, 2021, **594**(7862): 246-252
- [60] Watanabe Y, Allen J D, Wrapp D, *et al.* Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 2020, **369**(6501): 330-333
- [61] Tian W, Li D, Zhang N, *et al.* O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. *Cell Res*, 2021, **31**(10): 1123-1125
- [62] Dong X, Chen C, Yan J, *et al.* Comprehensive O-glycosylation analysis of the SARS-CoV-2 spike protein with biomimetic Trp-Arg materials. *Anal Chem*, 2021, **93**(30): 10444-10452
- [63] Wang P, Luo R, Zhang M, *et al.* A cross-talk between epithelium and endothelium mediates human alveolar-capillary injury during SARS-CoV-2 infection. *Cell Death Dis*, 2020, **11**(12): 1042
- [64] Li J, Guo M, Tian X, *et al.* Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med (NY)*, 2021, **2**(1): 99-112
- [65] Chen Y, Yao H, Zhang N, *et al.* Proteomic analysis identifies prolonged disturbances in pathways related to cholesterol metabolism and myocardium function in the COVID-19 recovery stage. *J Proteome Res*, 2021, **20**(7): 3463-3474
- [66] Harrison A G, Lin T, Wang P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends Immunol*, 2020, **41**(12): 1100-1115
- [67] Whetton A D, Preston G W, Abubeker S, *et al.* Proteomics and informatics for understanding phases and identifying biomarkers in COVID-19 disease. *J Proteome Res*, 2020, **19**(11): 4219-4232
- [68] D'alessandro A, Thomas T, Dzieciatkowska M, *et al.* Serum proteomics in COVID-19 patients: altered coagulation and complement status as a function of IL-6 level. *J Proteome Res*, 2020, **19**(11): 4417-4427
- [69] Matthay M A, Aldrich J M, Gotts J E. Treatment for severe acute respiratory distress syndrome from COVID-19. *Lancet Respir Med*, 2020, **8**(5): 433-434
- [70] Gordon D E, Jang G M, Bouhaddou M, *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 2020, **583**(7816): 459-468

Progress in Research of COVID-19 Based on Proteomics Techniques*

LIU Xiang^{1,2}, GONG Peng-Yun^{1,2}, TANG Min^{1,2}, HU Hong-Ke^{1,2},
ZHANG Yong-Biao^{1,2}** , LIU Chao^{1,2}**

⁽¹⁾*School of Engineering Medicine, Beihang University, Beijing 100191, China;*

⁽²⁾*School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China)*

Abstract Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This virus is highly contagious and spreads rapidly, posing a serious threat to the health and safety of people around the world. Proteomics technology has the characteristics of high throughput and high sensitivity, and plays an important role in biomarker discovery, molecular mechanism research, and therapeutic target research. Proteomics technology has been widely used in the research of COVID-19. Herein, this study provides a comprehensive review of the research progress of COVID-19 based on proteomics techniques. In section 1, the genome structure of SARS-CoV-2 and the process of SARS-CoV-2 infecting host cells were summarized. In section 2, the currently commonly used mass spectrometry (MS)-based proteomics techniques including liquid chromatography (LC)-MS and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS were reviewed. In section 3, the application progress in the research of precise diagnosis, molecular mechanism and drug therapy targets of COVID-19 based proteomics technology was highlighted. Proteomics have been employed in biomedical research to uncover biomarkers associated with COVID-19, it also provides a comprehensive snapshot of virus-induced changes to the host following infection, invasion, persistence, and pathogenesis and can prime the identification of novel therapeutic targets for preventing or lessening disease severity. In section 4, the future development direction of proteomics was prospected. It's hoped that this review can help to promote the development of proteomics technology in the precise diagnosis and treatment of diseases.

Key words proteomics, mass spectrometry, coronavirus disease 2019, biomarkers

DOI: 10.16476/j.pibb.2022.0029

* This work was supported by grants from The National Natural Science Foundation of China (32171442) and National Key Research and Development Program of China (2021YFA1301602, 2021YFA1301603).

** Corresponding author.

LIU Chao. Tel: 86-10-82338500, E-mail: liuchaobuaa@buaa.edu.cn

ZHANG Yong-Biao. Tel: 86-10-82338500, E-mail: Zhangyongbiao@buaa.edu.cn

Received: January 21, 2022 Accepted: April 19, 2022