

王洪博士, 中国医学科学院血液病医院(中国医学科学院血液学研究所), 实验血液学国家重点实验室和国家血液系统疾病临床医学研究中心, 细胞生态海河实验室研究员, 蛋白质组学平台负责人。长期致力于基于质谱的前沿蛋白质组学和代谢组学技术的开发及其在血液学领域的应用。近五年的主要研究成果在 *Cancer Cell*、*Nature Communications*、*Genome Biology*、*Molecular Neurodegeneration* 和 *Clinical Proteomics* 等国际学术期刊上发表。

[https://www.x-mol.com/groups/Wang\\_Hong/publications](https://www.x-mol.com/groups/Wang_Hong/publications)

## 基于质谱的血液蛋白质组学: 血液学研究的新焦点

付莉霞<sup>#</sup> 程子倩<sup>#</sup> 王洪<sup>\*</sup> 牛明明<sup>\*</sup>

(中国医学科学院血液病医院(中国医学科学院血液学研究所), 实验血液学国家重点实验室, 国家血液系统疾病临床医学研究中心, 细胞生态海河实验室, 天津 300020)

**摘要** 随着包括单细胞蛋白质组学在内的前沿蛋白质组学技术日趋成熟, 其在血液学领域的应用也迅速扩展。该文简要介绍了基于质谱的蛋白质组学技术及其研究方法, 讨论了血液蛋白质组学领域的最新研究进展, 并对血浆/血清及外泌体蛋白质组学进行了分析与评述。最后对血液蛋白质组学的未来发展趋势进行了展望, 预计临床血液蛋白质组学必将成为血液学下个十年的研究焦点之一。

**关键词** 质谱; 蛋白质组学; 单细胞蛋白质组学; 血液学; 血液系统疾病; 血浆; 血清; 外泌体

## MS-Based Blood Proteomics: Emerging Research Focus in Hematology

FU Lixia<sup>#</sup>, CHENG Ziqian<sup>#</sup>, WANG Hong<sup>\*</sup>, NIU Mingming<sup>\*</sup>

(State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China)

**Abstract** With cutting-edge MS (mass spectrometry)-based proteomic technologies (e.g. single-cell proteomics) getting mature nowadays, their application in hematology is quickly expanding. This article briefly introduces MS-based proteomics, then reviews recent progress in blood proteomics, and further comments on the technological improvements of plasma/serum proteomics and the potential of extracellular vesicle proteomics. Finally, the future development of clinical blood proteome is prospected. Clinical blood proteomics will certainly become a research focus in near future.

收稿日期: 2021-11-05

接受日期: 2021-12-06

<sup>#</sup>共同第一作者

<sup>\*</sup>通讯作者。Tel: 022-23909114, E-mail: wanghong@ihcams.ac.cn; Tel: 022-23909287, E-mail: niumingming@ihcams.ac.cn

Received: November 5, 2021 Accepted: December 6, 2021

<sup>#</sup>These authors contributed equally to this work

<sup>\*</sup>Corresponding authors. Tel: +86-22-23909114, E-mail: wanghong@ihcams.ac.cn; Tel: +86-22-23909287, E-mail: niumingming@ihcams.ac.cn

**Keywords** mass spectrometry; proteomics; single-cell proteomics; hematology; blood diseases; plasma; serum; extracellular vesicles

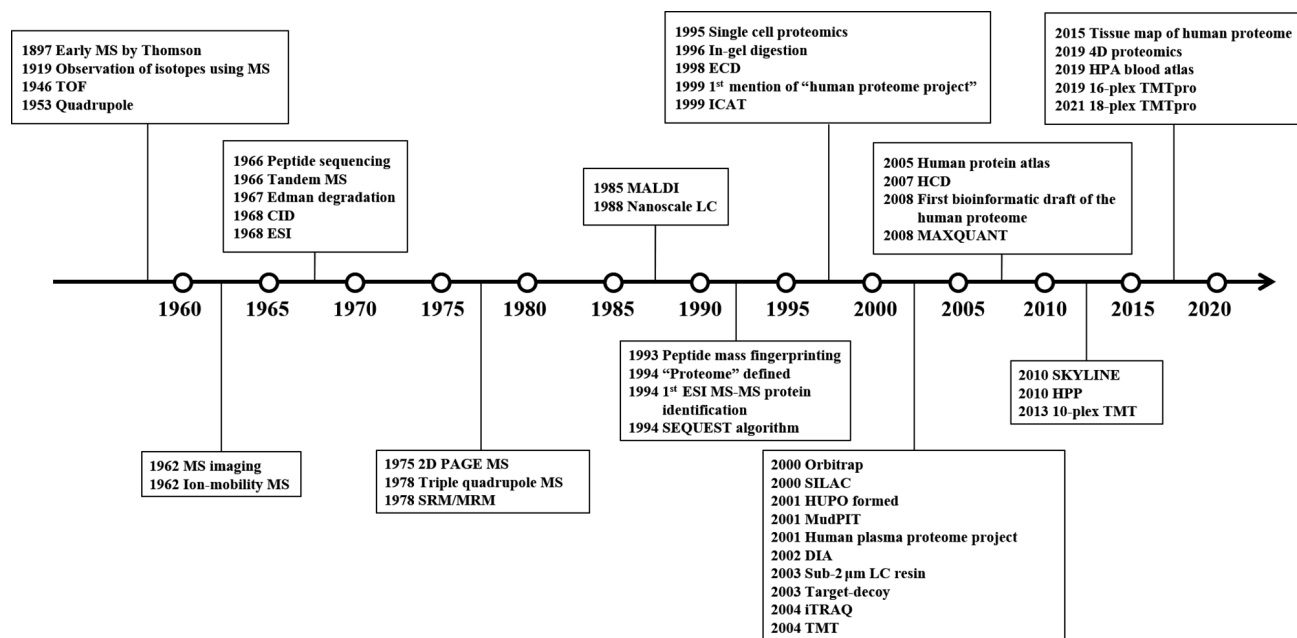
由中国医学科学院血液病医院(中国医学科学院血液学研究所)发起的血细胞图谱计划目前已取得多项重要研究成果,特别是基于单细胞转录组绘制的生理和病理情况下的血细胞图谱,为深入研究造血谱系的分化、干细胞的发育、各种血液疾病的发生发展机制及治疗方法建立了基础。绘制血细胞的蛋白质组学图谱将是血液学下一阶段的研究重点。随着蛋白质组学技术的飞速发展,特别是高通量和超灵敏的深度蛋白质组学技术及单细胞蛋白质组学技术的日益成熟,过去无法实现的许多临床血液蛋白质组学研究在今天已经成为了可能。比如,最新的超灵敏蛋白质组学技术已经实现了对纳克级微量样本的深度分析,为解析包括人类造血干细胞在内的各种微量临床样本的蛋白质组学图谱奠定了基础;高通量单细胞蛋白质组学技术的快速发展将实现在蛋白质水平精准解析造血谱系的

分化过程。除此之外,包括血浆/血清以及外泌体蛋白质组学等的关键技术突破也将把血液学的研究拓展至新的维度。

## 1 基于质谱的蛋白质组学技术的发展

### 1.1 基于质谱的蛋白质组学技术

1.1.1 简要历史回顾 质谱(mass spectrometry, MS)技术早在19世纪末就已经出现,但直到20世纪后半叶才逐渐应用于对蛋白质的研究。相较于1920年提出的基因组学,蛋白质组学的概念一直到1994年才被正式提出(图1)。1897年,THOMSON发现电子,随后发现带电离子在电场和磁场中的运动规律,为质谱技术作出了开创性的工作<sup>[1]</sup>。1966年,质谱技术被用于寡肽测序,成为质谱研究蛋白质的开端<sup>[2]</sup>。由于蛋白质大分子不易离子化,质谱技术一直局限于分子量小的耐热化合物研究。直到20世纪80年代末,



时间线展示从1960年前到现在质谱和蛋白质组学技术领域的代表性技术发展。CID: 碰撞诱导解离; SRM/MRM: 选择/多重反应监测; ECD: 电子捕获解离; ICAT: 同位素编码亲和和标签; SILAC: 细胞培养中稳定同位素标记; HUPO: 人类蛋白质组组织; iTRAQ: 相对和绝对定量等压标签; HCD: 高能碰撞解离; HPP: 人类蛋白质组项目; HPA: 人类蛋白质组图谱。

Timeline indicates representative technological development in the field of mass spectrometry and proteomics occurred from before 1960 to the present. CID: collision-induced dissociation; SRM/MRM: selected/multiple reaction monitoring; ECD: electron capture dissociation; ICAT: isotope-coded affinity tags; SILAC: stable isotope labelled amino acid for cell culture; HUPO: human proteome organization; iTRAQ: isobaric tags for relative and absolute quantification; HCD: higher-energy collisional dissociation; HPP: human proteome project; HPA: human protein atlas.

图1 质谱和蛋白质组学技术发展的简要历史回顾

Fig.1 Representative landmark events in mass spectrometry and proteomics

两种软电离方法——电喷雾电离(electrospray ionization, ESI)和基质辅助激光解吸电离(matrix assisted laser desorption ionization, MALDI)的开发才实现了生物大分子的温和离子化<sup>[3-4]</sup>。此后,质谱技术在生命科学领域得到广泛应用并迅速发展,如今已成为分子和细胞生物学研究的重要工具以及蛋白质组学领域的核心技术。

**1.1.2 质谱技术** 质谱是一种将化学物质电离为带电离子进而检测其质荷比的分析技术。质谱仪主要由离子源、质量分析器和检测器构成。离子源电离分析样品,将其转换为气相的带电离子,常用的离子源是ESI和MALDI。由于MALDI与液相分离方法不兼容,ESI目前已成为蛋白质组学分析中最常用的电离方法<sup>[5]</sup>。质量分析器通过电场或磁场控制离子的运动,确定分析物的质荷比(mass-to-charge ratio,  $m/z$ )及其相对强度,从而得到质谱图。早期常用的分析器有飞行时间(time of flight, TOF)、四极杆(quadrupole)、离子阱(ion trap)和线性离子阱(linear ion trap)等<sup>[6]</sup>。20世纪90年代末诞生的轨道阱(orbitrap)克服了早期设备在分辨率、质量精度、扫描速度及灵敏度等方面的不足,在蛋白质组学应用中显示出巨大的潜力<sup>[7]</sup>。目前,为应对生物样本中复杂的蛋白质组学成分,多通过结合不同分析器的特点,采用多种质量分析器串接的杂合质谱仪,如常用的Exploris 480和tims TOF pro等。蛋白质分析还可应用非质谱技术,如基于特异亲和试剂的方法(蛋白芯片<sup>[8]</sup>、ELISA和免疫组织化学)或者基于核酸适配体的方法,但这些方法过度依赖于试剂的特异性,不适用于复杂生物样本的大规模和深度分析<sup>[9]</sup>。伴随着质谱技术的飞速发展,基于质谱的蛋白质组学已逐渐成为后基因时代的学科焦点。

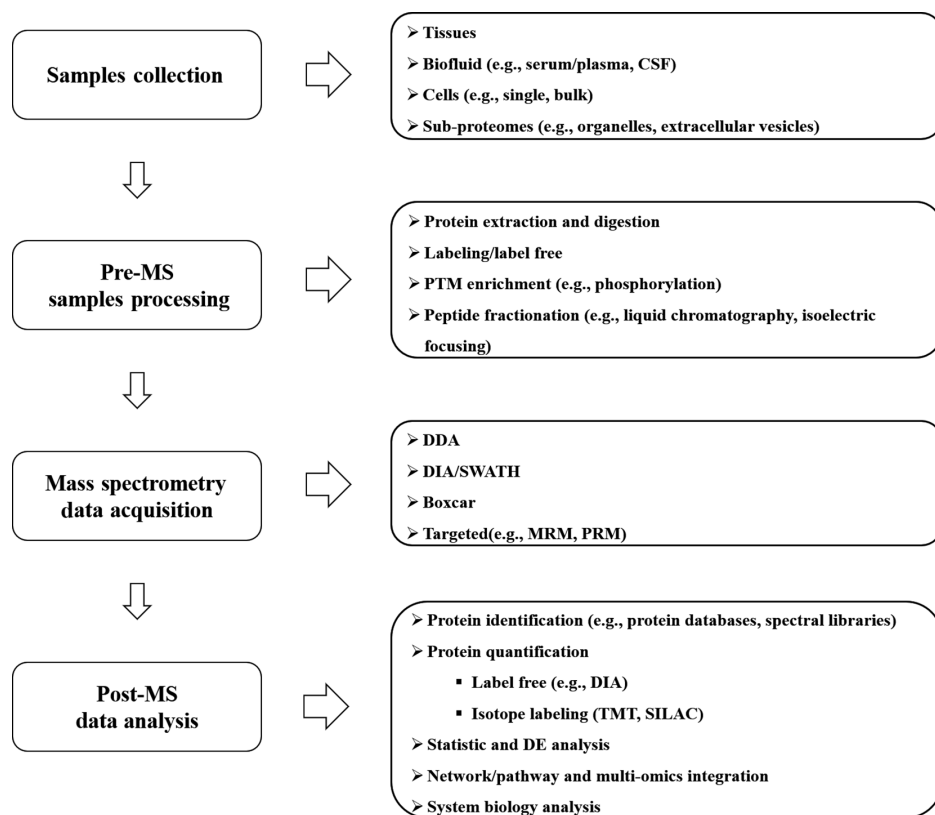
## 1.2 蛋白质组学分析的策略

**1.2.1 蛋白质组学分析的基本策略** 基于质谱的蛋白质组学分析策略通常包括:直接分析完整蛋白质的自上而下蛋白质组学(top-down proteomics)和对蛋白质消化后产生的肽段进行分析的自下而上蛋白质组学(bottom-up proteomics)。由于生物样本蛋白质组的高度复杂性,对完整蛋白质进行分离和质谱分析具有很大的挑战。相对而言,肽段更容易被分离、离子化和解离,因此bottom-up蛋白质组学的应用更为普遍,是目前对复杂蛋白质样本进行大规模和高通量分析的标准方法<sup>[10]</sup>。

**1.2.2 蛋白质组学分析步骤** 基于质谱的蛋白质组学分析主要包括3个步骤:样本预处理、质谱数据采集和生物信息学数据分析(图2)。在bottom-up蛋白质组学研究中,从不同样本中提取的蛋白质由蛋白质水解酶(如胰蛋白酶)酶切成肽段,肽段再经液相色谱(liquid chromatography, LC)进行分离,经由ESI进入质谱仪进行分析,产出具有质荷比信息的质谱谱图,再由蛋白质组学软件如MaxQuant、SEQUEST和JUMP等完成对质谱数据的蛋白质数据库搜索,从而完成对蛋白质的鉴定和接下来的定量分析<sup>[11-13]</sup>。在此基础上,再通过生物信息学方法找到差异蛋白,并将蛋白质组学数据与其他组学数据进行整合分析,以实现提出或者验证生物学假说的目的。

**1.2.3 蛋白质组学分析方法** 质谱数据采集主要有三种模式:数据依赖采集(data-dependent acquisition, DDA)、数据非依赖采集(data-independent acquisition, DIA)和靶向蛋白质组学(targeted proteomics)。三种方法各有优劣,DDA扫描速度快且灵敏度高,在蛋白质组学研究领域被广泛应用,但其缺点是在不同样本的对比分析中会出现缺失值<sup>[14]</sup>。DIA改善了DDA出现缺失值的问题,但其数据更复杂,因此数据分析较DDA繁杂,且需要在分析新样本前预先生成特定谱库。同时,由于DIA与串联质量标签(tandem mass tag, TMT)等化学标记不兼容,所以不适合进行样本的预分馏,因此鉴定的覆盖率和深度通常不高<sup>[15]</sup>。Targeted模式中的目标肽及其裂解特性是已知的,适合用于验证目标肽的存在或定量分析,而不能用于发现新的蛋白<sup>[16]</sup>。

**1.2.4 蛋白质组学的定量方法** 对不同生理病理状态下蛋白质的差异性表达进行精准定量分析,是蛋白质组学在生物医学研究中的关键。串联质谱技术(MS/MS)中,首先通过一级质谱(MS1)分析前体离子(比如蛋白质裂解产生的肽段),随后前体离子经碰撞裂解产生碎片离子,经由二级质谱(MS2)进行分析。常用的定量方法有无标记法(label-free)和基于化学标记法(labeling)。无标记的方法已经从半定量计数发展到基于MS1或MS2的更精准的蛋白定量分析。化学标记方法包括基于MS1(如ICAT和SILAC)和基于MS2(如iTRAQ、TMT和DiLeu等)的标记方法<sup>[17]</sup>。基于MS1的标记方法由于光谱更为复杂,因此鉴定效率较低,且最多只能标记2~3个样本;而基于MS2的标记方法则



基于质谱的蛋白质组学分析流程主要包括样本采集、样本预处理、质谱数据采集和生物信息学数据分析。CSF: 脑脊液; PTM: 蛋白质翻译后修饰; PRM: 平行反应监测。

Major steps in MS-based proteomics consist of sample collection, pre-MS sample processing, mass spectrometry data acquisition, and post-MS data analysis. CSF: cerebrospinal fluid; PTM: post-translational modification; PRM: parallel reaction monitoring.

图2 基于质谱的蛋白质组学分析流程

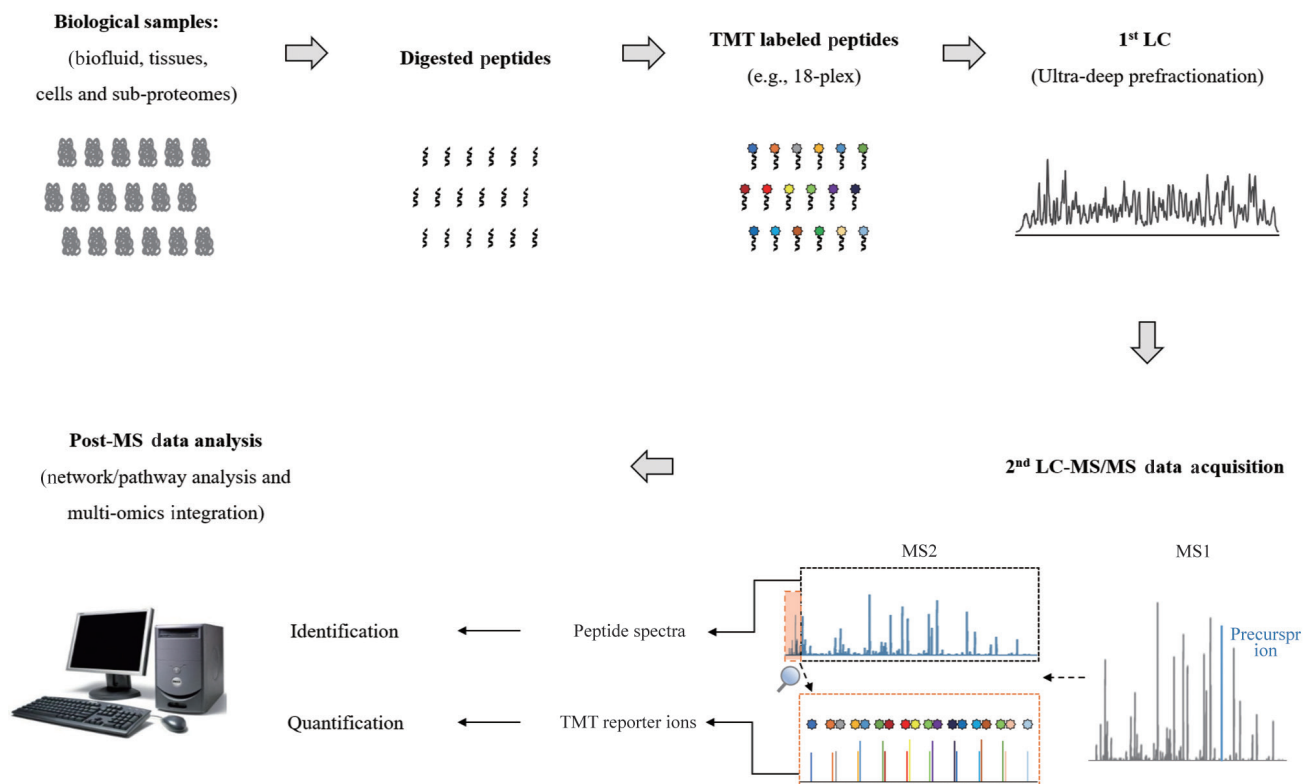
Fig.2 Schematic diagram of major steps in a standard MS-based proteomics analysis

极大地提高了分析的通量, 如最新的TMT试剂已经能实现同时对18个样本进行标记。TMT标记法由于其高复用能力和高蛋白质组覆盖率已被广泛用于大规模蛋白质组学研究之中<sup>[18]</sup>, 但基于MS2的标记方法也存在高噪音干扰引起的比值压缩(ratio compression)问题。我们通过优化分析策略并结合质谱后校正等方法, 有效地降低了干扰影响, 显著地提高了定量的精准度<sup>[19]</sup>。

**1.2.5 蛋白质的翻译后修饰** 蛋白质可以通过各种翻译后修饰来增加其功能的多样性。人体细胞通过各种蛋白质翻译后修饰, 动态调节着几乎所有的细胞活动和生物功能。常见的修饰类型包括磷酸化、泛素化及糖基化等<sup>[20]</sup>。基于质谱的蛋白质组学方法同样适用于翻译后修饰的大规模分析。由于样本中发生修饰的肽段丰度低且动态范围广, 质谱检测前常需要对修饰蛋白进行富集, 如使用金属离子(IMAC及TiO<sub>2</sub>等)对磷酸化蛋白进行富集或采用抗

体或凝集素等对其他多种修饰蛋白进行富集<sup>[21]</sup>。目前, 最新的质谱技术已经能够对数万个磷酸化位点进行高通量检测和精准定量<sup>[22]</sup>。

**1.2.6 高通量深度蛋白质组学** 蛋白质鉴定的深度和覆盖率是评估蛋白质组学技术的重要指标。在生理病理过程中发挥关键调控作用的蛋白质丰度通常较低, 因此需要更高深度的蛋白质组学平台对其进行分析。目前应用较广的是与TMT串联标记相结合的液相色谱质谱联用平台——TMT-LC/LC-MS/MS(图3)<sup>[23]</sup>。该平台充分利用了TMT标记法在提高样本通量上的优势, 将其与高分辨率二维液相色谱和串联质谱联用, 成功实现了同时在27个生物样本中对10 000个左右蛋白质的精准定量分析<sup>[18]</sup>。目前应用较广的深度蛋白质组学研究方法还包括液相色谱和离子迁移质谱联用的LC-IMS-DIA-MS平台, 该平台也可实现对组织蛋白的深度分析<sup>[24]</sup>, 但是由于其与TMT标记法不兼容, 且不适合联用二维液相色



流程图展示了基于TMT-LC/LC-MS/MS的高通量深度蛋白质组学平台的主要实验步骤。

Schematic workflow shows the major steps involved in TMT-LC/LC-MS/MS-based high-throughput deep proteomics platform.

图3 代表性高通量深度蛋白质组学平台的流程图

Fig.3 Schematic workflow of a MS-based high-throughput deep proteomic pipeline

谱进行样本预分馏,因此在蛋白质组的检测深度上不及TMT-LC/LC-MS/MS平台。

作为后基因组时代的学科焦点,目前的蛋白质组学及其相关技术正在蓬勃发展,人体蛋白质组图谱草案已经报道了92%蛋白质编码序列的质谱证据<sup>[25]</sup>。最新的高通量、高灵敏及高深度的蛋白质组学平台已经能够实现在小于微克级的临床组织或者细胞等复杂生物样本中快速地定量分析上万个蛋白质<sup>[26]</sup>。基于质谱的蛋白质组学技术已经成为生物医学研究中不可或缺的部分。

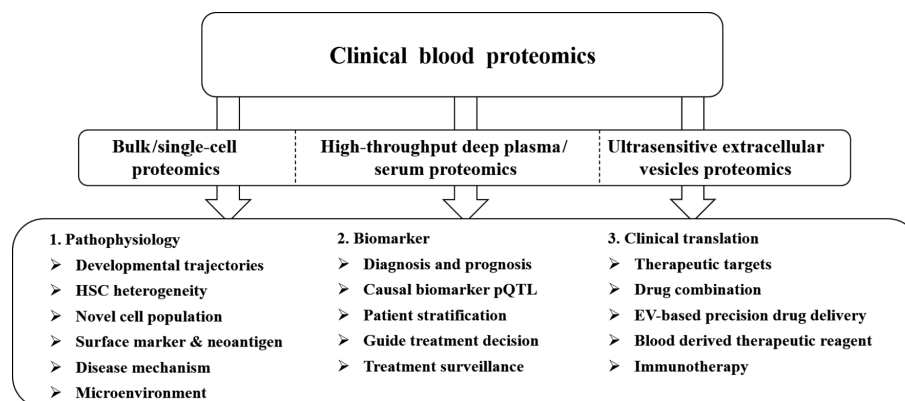
## 2 基于质谱的蛋白组学在血液系统中的应用

### 2.1 基于质谱的蛋白组学在血细胞研究中的应用

蛋白质是基因密码的效应者,是所有细胞功能的执行分子,同时也是绝大多数药物的作用靶点。尽管全基因组测序(whole genome sequencing, WGS)、全外显子测序(whole exome sequencing, WES)及RNA测序(RNA-sequencing, RNA-Seq)已经

从基因层面极大地提升了我们对血液学和血液疾病的认知。但有研究表明,在血细胞中mRNA和蛋白质表达水平的相关性较差,且许多蛋白质翻译后修饰,如蛋白质的磷酸化及泛素化等,在调节血细胞生物功能中起着至关重要的作用<sup>[27]</sup>。因此,对血细胞蛋白质组和翻译后修饰组的深度解析能够填补血液学领域的许多空白,完善对各种血液系统疾病的理解。基于质谱的蛋白质组学技术已经成为所有涉及蛋白质的生物学研究的重要支柱,被广泛应用于所有生物医学研究,在阐明疾病发生发展机制和探索治疗靶点等领域取得重要成果的同时,在对各种血细胞的研究中也取得了诸多令人欣喜的成果(图4)。

2.1.1 基于质谱的蛋白质组学与血液系统疾病机制研究 基于质谱的蛋白质组学在探索血液疾病机制的研究中发挥着关键的作用。RAFFEL等<sup>[28]</sup>通过蛋白质组学研究证实了,与正常的造血干祖细胞相比,白血病干细胞有自己独特的代谢特征和靶向通路,且其代谢途径的改变主要表现在蛋白质组而非转录组水平。SHI等<sup>[29]</sup>利用蛋白质相互作用组筛查,



基于质谱的新兴蛋白质组学技术在前沿血液学和血液病的基础、转化和临床研究中的应用。HSC: 造血干细胞; pQTL: 蛋白质数量性状位点。Cutting-edge MS-based proteomics technologies to address emerging basic, translational and clinical research questions in the hematology and blood diseases field. HSC: hematopoietic stem cell; pQTL: protein quantitative trait loc.

图4 基于质谱的蛋白质组学在血液学和血液系统疾病研究中的应用

Fig.4 Application of MS-based proteomics in hematology and blood diseases

首次发现了氨基酸调控免疫细胞功能的机制,证实了氨基酸信号转导是允许和维持 mTORC1 激活和 Treg 细胞功能编程的关键因素。DEEB 等<sup>[30]</sup>通过对弥漫性大 B 细胞淋巴瘤患者的冰冻组织样本进行基于 SILAC 定量的蛋白质组学分析,发现了部分在无进展患者中特有的蛋白质的表达,为临床患者亚型分类提供了蛋白质层面的证据。MCDONNELL 等<sup>[31]</sup>还通过整合蛋白质磷酸化组和代谢组,发现了其特有的磷酸化通路及代谢重编程,揭示了间变性大淋巴瘤患者的发病机制。此外,通过对大队列慢性淋巴细胞白血病 (chronic lymphocytic leukemia, CLL) 患者的临床样本分析,MEIER-ABT 等<sup>[32]</sup>揭示了 CLL 的蛋白质组学图谱,并解释了 CLL 靶向药物的作用机制。我们与合作者通过整合深度蛋白质组和磷酸化组的网络分析,系统地揭示了调控 T 细胞激活的网络信号通路和关键调控因子<sup>[33]</sup>。

**2.1.2 明确不同谱系造血细胞发育途径和表面蛋白表达特征及其功能** 对处于不同发育阶段的血细胞的蛋白质组学解析,是从分子层面系统阐明造血细胞发育过程、理解由发育异常导致的各种血液疾病和探索新的血细胞亚型的关键。XU 等<sup>[34]</sup>通过深度蛋白质组学筛查,发现了调节红细胞发育的关键调控因子和新的调控机制,同时证明了对蛋白质降解的调控在红细胞发育过程中的重要作用。通过我们开发的深度蛋白质组和磷酸化组分析平台,KENNEDY 等<sup>[35]</sup>成功发现了新的免疫细胞亚群,并重新剖析了 B 细胞的表型及功能。此外,UNWIN 等<sup>[36]</sup>还

通过蛋白质组学筛查,发现了干细胞及其后代的分子特征,并明确了干细胞自我更新和骨髓衰竭的机制。基于质谱的蛋白质组学能够进一步解析不同谱系细胞的表面蛋白、发育过程及其机制,并有助于发现新的细胞亚群,为阐明血液系统疾病的发生发展提供理论依据。

**2.1.3 发现血液系统疾病治疗靶点** 基于质谱的蛋白质组学和以蛋白质组学为中心的多组学整合研究已在探索各种疾病精准靶向治疗方法的研究中展现出了巨大的潜力和应用前景。通过应用我们最新的全蛋白质组和磷酸化组分析平台并结合单细胞转录组以及先进的网络分析方法,GOCHO 等<sup>[37]</sup>证实了在 T 细胞急性淋巴细胞白血病 (T-lineage acute lymphoblastic leukemia, T-ALL) 患者中 LCK 和 BCL2 信号转导以及 T 细胞在成熟阶段的异质性,从而解释了其靶向治疗的作用机制及癌细胞对药物的敏感性。此外,我们与合作团队通过超敏细胞膜蛋白质组与转录组等多组学分析,发现了位于小儿 T-ALL 癌细胞表面的新特异性抗原,并生成了新的 CAR-T 细胞用于治疗,取得了良好的治疗效果,该研究证明了超灵敏细胞表面蛋白质组分析在探索新的免疫治疗靶点的巨大潜力<sup>[38]</sup>。再者,通过蛋白质组学和磷酸化组学可以明确白血病细胞系对单靶点药物的耐药机制,为指导临床用药提供理论依据<sup>[39]</sup>。基于质谱的蛋白质组学研究能够发现无法被转录组检测到的血液系统疾病的潜在药物靶点,并与其他组学研究结合,明确耐药机制,进而为临床病人的个体化用药提

供指导。

目前基于质谱的蛋白质组学已经被用于在血细胞的发育分化、血液系统疾病的发生发展机制和靶向用药等研究中进行诸多探索并取得重要成果。相信随着各种前沿质谱技术的不断成熟,蛋白质组学将在血细胞和各种血液病的研究中发挥愈加重要的作用。

## 2.2 血浆/血清蛋白质组

根据WHO疾病分类组织(international classification of disease, ICD)的统计,人类疾病的种类超过14 500种,然而血浆/血清作为临床检验最重要的样本,其来源的蛋白质生物标记物目前仅有不足200个被FDA批准用于临床使用。因此,血浆/血清蛋白质组学领域亟需先进的技术来探索新的血浆/血清蛋白质生物标志物。虽然血细胞只占全血的45%,血浆占超过一半,但是相较于在血细胞领域的广泛应用和所取得的重要成果,对血浆/血清的蛋白质组学研究所取得的成果则非常局限,这主要是因为其极端的复杂性和超广的蛋白质丰度范围,如丰度最高的20种蛋白质就占血浆蛋白质组总量的90%以上,而蛋白质组丰度跨度更是超过10个数量级,在血浆里检测低丰度蛋白质的难度就如同在地球上所有人口里准确定位到某个个体,这为血浆/血清蛋白的研究带来极大的挑战。不过,血浆/血清蛋白质组学技术已实现了不少突破,随着新技术的不断发展,我们相信基于血浆/血清蛋白质组学的研究将在探索疾病新的生物标志物方面实现更多突破。

**2.2.1 血浆/血清蛋白质组学的研究瓶颈** 筛查深度的不足是限制质谱技术在血浆/血清中发现新的疾病标志物的主要瓶颈之一。目前绝大多数基于质谱的筛查技术在血浆/血清中只能检测到不超过500个的丰度最高的蛋白质,在血浆/血清蛋白质组中的覆盖率不足5%,而目前已被批准临床使用的标志物绝大多数都来自于这500个蛋白质,这意味着占血浆蛋白质总数95%以上的蛋白质仍没法被当前绝大多数的筛查技术检测到。虽然目前已有部分方法可以通过去除高丰度血浆蛋白质来提高检测深度,但却存在高丰度蛋白质去除不完整、非特异性去除低丰度蛋白质以及实验误差大等不足。除检测深度外,另一个关键难点是实验方法的通量,由于临床血液样本的个体差异较大,血浆生物标志物的研究通常至少需要分析数百个临床样本。因此,提高分析临

床样本的通量也是目前血浆/血清蛋白质组学技术的主要发展方向。

**2.2.2 血浆/血清蛋白质组学的研究进展** 为了提高蛋白质组学筛查的深度,目前常用的研究手段包括通过免疫耗竭预先去除血浆中的高丰度蛋白质和通过液相色谱等进行复杂蛋白质组的预分馏以降低高丰度蛋白质对血浆中低丰度蛋白质的检测的干扰。此外多个课题组也在通过不断改进蛋白质组学分析方法来提高检测通量,GEYER等<sup>[40]</sup>通过结合自动化样本处理和快速液相色谱技术等实现了每天检测60个样本,大大提高了检测的通量,但却在一定程度上牺牲了检测深度,每个样本只能定量分析不到500个蛋白质。为攻克上述难点,我们开发了一个新的平台,实现了在保持较高检测通量的同时,将血浆蛋白质定量分析覆盖率从5%左右提高到了50%左右,该平台是目前已报道过的最深度的血浆/血清蛋白质组学平台<sup>[41]</sup>。该平台通过TMT标记将检测通量提高了十倍以上,同时免除了免疫耗竭的步骤,并通过二维液相色谱对复杂肽段混合物进行深度的预分离,首次实现了在单次实验中同时定量分析11个血浆样本中接近5 000个的血浆蛋白质。目前,我们最新的平台已经能够将单个实验中的样本分析数量从11个提高至29个。我们相信未来的检测平台将在检测通量和蛋白质覆盖率之间找到一个平衡,以同时实现高通量和高深度的血浆蛋白质组学研究。

目前,每年只有1到2个新的血浆蛋白质生物标志物获准临床使用,而绝大部分临床正在使用的疾病标志物在上个世纪之前就已经推出。尽管已经有许多通过基于质谱的血浆蛋白质组学筛查疾病生物标志物的研究,但却几乎均未能实现临床的转化应用,这主要是受限于过去技术在检测深度和通量上的缺陷。近年来,随着新技术在各方面的不断突破,我们有理由相信血浆蛋白质组学研究将为血液学领域的发展拓展出一片新的天地。

## 2.3 血细胞或血浆来源的外泌体蛋白质组

外泌体(extracellular vesicles, EVs)是当前生物医学领域的一个研究热点。外泌体是由细胞主动分泌到胞外的有膜包裹的微囊泡,是调控细胞间信息传递的关键介质,参与着几乎所有正常生理功能和疾病的发生发展,在疾病生物标志物、药物精准靶向递送和作为新的生物药剂等临床转化研究中展现出巨大的潜力。由血细胞和血浆分泌的外泌体通过

血液系统可以转移到人体所有器官和组织,能参与任何疾病的发生发展中。外泌体富含蛋白质,且相较于血液中很多因细胞损伤破裂而被动漏出到血液里的非功能性蛋白质,外泌体蛋白质通常具有重要生物学功能,因此具有更大的潜力成为疾病的特异性生物标志物。同时,与暴露在富含蛋白质水解酶的血液中的其他蛋白质不同,外泌体蛋白质被外膜包裹在内,不会被血液蛋白质水解酶降解,避免了血浆中高丰度蛋白质对外泌体蛋白质组研究的干扰。因此,对血细胞和血浆来源的外泌体蛋白质组的研究有望成为血液学领域的新兴热点之一。

**2.3.1 外泌体蛋白质组学目前的研究瓶颈** 分离出足够数量和纯度的外泌体是进行外泌体蛋白质组学研究和发现生物标志物的关键。尽管目前已有包括超速离心、过滤、尺寸排除层析、免疫亲和捕获和基于微芯片等大量针对外泌体分离纯化的方法,但由于外泌体纳米级的微小尺寸以及极低的蛋白质产量,高效的提取外泌体进行蛋白质组学分析仍然是一个较大的挑战。除此之外,EVs在生物发生、大小、物理特性、含量和功能等方面都存在高度的异质性,这也为其分离纯化带来了极大的困难。尽管目前还有许多挑战需要攻克,但是外泌体蛋白质组学研究所展现出的巨大潜力和应用前景值得血液学领域更多的关注和投入。

**2.3.2 外泌体蛋白质组学在血液学领域的研究进展** 外泌体蛋白质组学技术通常被用作探索临床疾病的潜在生物标志物。PRIETO等<sup>[42]</sup>通过血浆外泌体蛋白组学分析,在找到辅助诊断进展期CLL的潜在标志物的同时,证实了S100A9<sup>+</sup>外泌体可以促进CLL进展的特殊机理。中国医学科学院血液学研究所的程涛和程辉教授及其合作团队<sup>[43]</sup>通过蛋白质组学筛查首次证实了由血管内皮细胞分泌的ANGPTL2的外泌体能加速白血病的进程。对于血液系统疾病而言,外泌体蛋白质组学还可以用来阐明外泌体在血液系统中的生理病理功能、探索疾病发生发展机制以及开发基于外泌体的药物精准靶向载体和疾病治疗新策略。

总之,随着外泌体分离、定量和质控分析方法的逐渐成熟,以及基于质谱的超敏外泌体蛋白质组学技术的不断优化,对血液外泌体的深入探讨将为血液系统疾病的基础和临床转化研究提供新的思路,如开发应用于血液系统的外泌体药物载体以用

于疾病的精准靶向治疗、提高外泌体的递送药物效率和特异性以及开发造血干细胞的外泌体药物制剂等。

### 3 总结和展望

血液学研究的新阶段迫切需要系统地从蛋白质层面加深对血液系统疾病的理解,并与基因组、转录组和代谢组等多组学相结合以指导基础研究、临床应用和治疗创新,从而实现对疾病的精准临床诊疗。

随着二代测序技术的发展和血细胞分子图谱计划的提出,程涛教授领衔的多个研究组已经围绕血细胞发育、分化维持及病变等方面进行了一系列研究,包括但不限于绘制人和小鼠各类血细胞的单细胞转录组图谱、小鼠造血系统分化层级的RNA编辑图谱以及再生障碍性贫血病人的血液病理图谱等,从基因组和转录组层面注释了血细胞在生理和病理过程中的改变,为后续血液生理学和病理学研究提供了重要的依据和参考。血细胞图谱计划新阶段的研究重点是将这些重要学术问题的研究重心延展到蛋白质组学层面,绘制各种生理和病理情况下的血细胞、血浆/血清及其来源的外泌体的蛋白质组学图谱。

蛋白质组学技术在过去十年间的革命性进展和不断优化让期盼已久的对大规模临床样本的研究成为可能。由于临床样本具有个体化差异大以及样本微量等特点,临床蛋白质组学技术将继续朝着微量化、高深度和高通量的方向持续发展。除了色谱和质谱技术的持续全方位提升以外,样本处理的微量化、自动化、智能化和模块化将会得到更高的重视。基于人工智能的多组学网络分析方法和建模也将持续提高对大数据挖掘的效率。相信基于质谱的蛋白质组学技术将会迅速从实验室研究延展到临床应用。

单细胞分子图谱的构建与研究是近年来全球生物医药领域研究的前沿与热点。单细胞蛋白质组学技术也在近几年飞速发展,目前最新的单细胞蛋白质组学技术已经可以在较高通量的基础上实现在单个细胞中分析1 000个左右的蛋白质<sup>[44]</sup>,未来的单细胞技术将继续朝着更高通量、更深的蛋白质组检测深度和更可靠的定量分析方向持续发展。随着单细胞蛋白质组学技术的日趋成熟,相信其将在解析造



血细胞的发育轨迹和探索新的血细胞亚群等领域取得重大突破,革新我们对造血谱系的认知。

总之,技术的进步已经让曾经的许多愿景变为今天的现实,基于质谱的蛋白质组学研究已经在血液学领域取得了好的开端和多项突破。随着新技术的不断推出、研究的继续深入和投入的不断加大,血液蛋白质组学的未来值得期待。

### 参考文献 (References)

- [1] THOMSON J J. XXVI. Rays of positive electricity [J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2009, 21(122): 225-49.
- [2] BIEMANN K, TSUNAKAWA S, SONNENBICHLER J, et al. Structure of an odd nucleoside from serine-specific transfer ribonucleic acid [J]. Angew Chem Int Ed Engl, 1966, 5(6): 590-1.
- [3] FENN J B, MANN M, MENG C K, et al. Electrospray ionization for mass spectrometry of large biomolecules [J]. Science, 1989, 246(4926): 64-71.
- [4] TANAKA K, WAKI H, IDO Y, et al. Protein and polymer analyses up to  $m/z$  100 000 by laser ionization time-of-flight mass spectrometry [J]. Rapid Commun Mass Spectrom, 1988, 2(8): 151-3.
- [5] MANN M. The ever expanding scope of electrospray mass spectrometry—a 30 year journey [J]. Nat Commun, 2019, 10(1): 3744.
- [6] GRIFFITHS J. A brief history of mass spectrometry [J]. Anal Chem, 2008, 80(15): 5678-83.
- [7] ELIUK S, MAKAROV A. Evolution of orbitrap mass spectrometry instrumentation [J]. Annu Rev Anal Chem, 2015, 8: 61-80.
- [8] ZHU H, SNYDER M. Protein chip technology [J]. Curr Opin Chem Biol, 2003, 7(1): 55-63.
- [9] NIMJEE S M, WHITE R R, BECKER R C, et al. Aptamers as Therapeutics [J]. Annu Rev Pharmacol Toxicol, 2017, 57: 61-79.
- [10] HAN X, ASLANIAN A, YATES J R, 3RD. Mass spectrometry for proteomics [J]. Curr Opin Chem Biol, 2008, 12(5): 483-90.
- [11] ENG J K, MCCORMACK A L, YATES J R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database [J]. J Am Soc Mass Spectrom, 1994, 5(11): 976-89.
- [12] COX J, MANN M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification [J]. Nat Biotechnol, 2008, 26(12): 1367-72.
- [13] WANG X, LI Y, WU Z, et al. JUMP: a tag-based database search tool for peptide identification with high sensitivity and accuracy [J]. Mol Cell Proteomics, 2014, 13(12): 3663-73.
- [14] DEFOSSEZ E, BOURQUIN J, VON REUSS S, et al. Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics [J]. Mass Spectrom Rev, 2021, doi: 10.1002/mas.21715.
- [15] LUDWIG C, GILLET L, ROSENBERGER G, et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial [J]. Mol Syst Biol, 2018, 14(8): e8126.
- [16] DOERR A. Mass spectrometry-based targeted proteomics [J]. Nat Methods, 2013, 10(1): 23.
- [17] PAPPIREDDI N, MARTIN L, WÜHR M. A review on quantitative multiplexed proteomics [J]. ChemBioChem, 2019, 20(10): 1210-24.
- [18] WANG Z, YU K, TAN H, et al. 27-plex tandem mass tag mass spectrometry for profiling brain proteome in Alzheimer's disease [J]. Anal Chem, 2020, 92(10): 7162-70.
- [19] NIU M, CHO J H, KODALI K, et al. Extensive peptide fractionation and y(1) ion-based interference detection method for enabling accurate quantification by isobaric labeling and mass spectrometry [J]. Anal Chem, 2017, 89(5): 2956-63.
- [20] MANN M, JENSEN O N. Proteomic analysis of post-translational modifications [J]. Nat Biotechnol, 2003, 21(3): 255-61.
- [21] ZHAO Y, JENSEN O N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques [J]. Proteomics, 2009, 9(20): 4632-41.
- [22] STEWART E, MCEVOY J, WANG H, et al. Identification of therapeutic targets in rhabdomyosarcoma through integrated genomic, epigenomic, and proteomic analyses [J]. Cancer Cell, 2018, 34(3): 411-26.e9.
- [23] LIU D, YANG S, KAVDIA K, et al. Deep profiling of microgram-scale proteome by tandem mass tag mass spectrometry [J]. J Proteome Res, 2021, 20(1): 337-45.
- [24] MEIER F, BRUNNER A D, KOCH S, et al. Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer [J]. Mol Cell Proteomics, 2018, 17(12): 2534-45.
- [25] WILHELM M, SCHLEGL J, HAHNE H, et al. Mass-spectrometry-based draft of the human proteome [J]. Nature, 2014, 509(7502): 582-7.
- [26] AEBERSOLD R, MANN M. Mass-spectrometric exploration of proteome structure and function [J]. Nature, 2016, 537(7620): 347-55.
- [27] ZARO B W, NOH J J, MASCETTI V L, et al. Proteomic analysis of young and old mouse hematopoietic stem cells and their progenitors reveals post-transcriptional regulation in stem cells [J]. Elife, 2020, 9: e62210.
- [28] RAFFEL S, KLIMMECK D, FALCONE M, et al. Quantitative proteomics reveals specific metabolic features of acute myeloid leukemia stem cells [J]. Blood, 2020, 136(13): 1507-19.
- [29] SHI H, CHAPMAN N M, WEN J, et al. Amino acids license kinase mTORC1 activity and Treg cell function via small G proteins RAG and RHEB [J]. Immunity, 2019, 51(6): 1012-27.e7.
- [30] DEEB S J, D'SOUZA R C, COX J, et al. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles [J]. Mol Cell Proteomics, 2012, 11(5): 77-89.
- [31] MCDONNELL S R, HWANG S R, ROLLAND D, et al. Integrated phosphoproteomic and metabolomic profiling reveals NPM-ALK-mediated phosphorylation of PKM2 and metabolic reprogramming in anaplastic large cell lymphoma [J]. Blood, 2013, 122(6): 958-68.
- [32] MEIER-ABT F, LU J, CANNIZZARO E, et al. The protein landscape of chronic lymphocytic leukemia (CLL) [J]. Blood, 2021, doi: blood.2020009741.
- [33] TAN H, YANG K, LI Y, et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation [J]. Immu-

- nity, 2017, 46(3): 488-503.
- [34] XU P, SCOTT D C, XU B, et al. FBXO11-mediated proteolysis of BAHD1 relieves PRC2-dependent transcriptional repression in erythropoiesis [J]. *Blood*, 2021, 137(2): 155-67.
- [35] KENNEDY D E, OKOREEH M K, MAIENSCHIN-CLINE M, et al. Novel specialized cell state and spatial compartments within the germinal center [J]. *Nat Immunol*, 2020, 21(6): 660-70.
- [36] UNWIN R D, SMITH D L, BLINCO D, et al. Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells [J]. *Blood*, 2006, 107(12): 4687-94.
- [37] GOCHO Y, LIU J, HU J, et al. Network-based systems pharmacology reveals heterogeneity in LCK and BCL2 signaling and therapeutic sensitivity of T-cell acute lymphoblastic leukemia [J]. *Nat Cancer*, 2021, 2(3): 284-99.
- [38] HEBBAR N, QU C, WANG H, et al. Multi-omic based antigen discovery for the immunotherapy of pediatric acute T cell lymphoblastic leukemia [J]. *Blood*, 2020, 136(Supplement 1): 17-18.
- [39] JOSHI S K, NECHIPORUK T, BOTTOMLY D, et al. The AML microenvironment catalyzes a stepwise evolution to gilteritinib resistance [J]. *Cancer Cell*, 2021, 39(7): 999-1014.e8.
- [40] GEYER P E, AREND F M, DOLL S, et al. High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion [J]. *EMBO Mol Med*, 2021, 13(8): e14167.
- [41] DEY K K, WANG H, NIU M, et al. Deep undepleted human serum proteome profiling toward biomarker discovery for Alzheimer's disease [J]. *Clin Proteomics*, 2019, 16: 16.
- [42] PRIETO D, SOTELO N, SEIJA N, et al. S100-A9 protein in exosomes from chronic lymphocytic leukemia cells promotes NF- $\kappa$ B activity during disease progression [J]. *Blood*, 2017, 130(6): 777-88.
- [43] HUANG D, SUN G, HAO X, et al. ANGPTL2-containing small extracellular vesicles from vascular endothelial cells accelerate leukemia progression [J]. *J Clin Invest*, 2021, 131(1): e138986.
- [44] SLAVOV N. Single-cell protein analysis by mass spectrometry [J]. *Curr Opin Chem Biol*, 2021, 60: 1-9.