

生物质谱分析新方法开发和多组学实验室建设

高妍, 张向阳

(天津大学药物科学与技术学院, 天津 300072)

摘要: 临床样本的蛋白质组检测面临被测物动态范围大、分析物浓度低、基质效应影响大等多方面的巨大挑战, 因此对蛋白质谱分析的灵敏度、动态范围、定量准确性和精密度都提出较高要求。团队建立了基于数据非依赖性采集 (Data Independent Acquisition, DIA) 质谱技术的高通量蛋白定量分析流程, 并不断优化生物质谱分析相应的样品处理方法和质谱检测参数, 针对不同科研需求设计实验流程。同时, 建立针对代谢组学、脂质组学、质谱成像和质谱流式的测试方法, 这些方法的开发和标准操作规程的确立为生命科学研究及临床医学研究提供了强有力的工具。

关键词: 生物质谱精准医学多组学研究 DIA 技术方法开发

中图分类号: Q26

文献标识码: A

在精准医学和临床研究中, 需要精确寻找到疾病的原因和治疗的靶点, 对患者分子生物学特征做出精准判断, 以最终实现对于疾病和特定患者个体化精准治疗的目的^[1-3]。目前被广泛接受的临床质谱应用领域主要是针对小分子化合物, 而蛋白质谱的临床应用还处于探索阶段^[4-6]。临床样本的复杂性, 使得蛋白质定性定量分析充满了挑战, 生物质谱由于其高分辨率和高准确度, 在靶点发现、精准分子分型、药物作用机制和药效方面都有着巨大的潜力。因此, 急需开发深度覆盖、定量结果准确、快速高通量的分析方法^[7]。团队以蛋白质谱分析技术为核心, 结合多种前沿质谱技术, 包括 DIA 技术、平行反应监测 (Parallel Reaction Monitoring, PRM) 靶向质谱分析、质谱流式 (Mass Cytometry) 单细胞蛋白质检测技术、以及基于离子淌度的 4D 蛋白质组学等, 并整合代谢组学、脂质组学和质谱成像等其它高通量组学技术, 建立了多组学临床质谱分析技术平台。

与数据依赖性采集 (Data Dependent Acquisition, DDA) 不同, DIA 分析技术因其覆盖度高、信号无损失、定量能力强的特点, 在临床蛋白质谱上有很好的应用前景。经典的 DIA 分析流程, 需要采用分馏样品的 DDA 检测数据建立谱图库以实现蛋白鉴定, 但这对临床微量蛋白样品来说难以实现^[8]。本文探讨了基于谱图库预测策略的 DIA 蛋白质组学分析流程, 如图 1 所示, 实现了微量蛋白质组学的深度覆盖和准确定量检测。该方法为临床微量样本检测和精准医疗开展提供了方法支撑。



图 1 基于谱图库预测策略的 DIA 的生物质谱分析方法流程图

1 样本收集

样本可为细胞、组织、体液等，本工作中以 HEK293T 细胞为研究对象。

2 样品前处理

本方法采用基于超滤管的制备方法（filter-aided sample preparation, FASP）^[9]。

2.1 裂解

取少量 293T 细胞，加入含 8M 尿素、cocktail 和 50mM 碳酸氢铵的裂解液，非接触式超声 8-10 个循环。

2.2 测蛋白浓度

离心，取上清，用 BCA 法测定蛋白浓度。

还原：取一定量蛋白，加入二硫苏糖醇 DTT，使其终浓度为 10mM，室温反应 45 分钟。

2.3 烷基化

加入碘乙酰胺 IAA，使其终浓度为 40mM，避光反应 45 分钟。

2.4 终止

加入二硫苏糖醇 DTT，使其终浓度为 40mM，室温反应 30 分钟。

2.5 转移

将蛋白转移到 10KD 超滤管中，14000g，离心 30 分钟，并用 50mM 碳酸氢铵洗两次，以除去所有盐类。

2.6 酶解

按照蛋白和酶的比例 25:1 加入胰蛋白酶，37°C 过夜反应 12-16 小时。

2.7 收集

换上新的接收管，14000g，离心 30 分钟，再用 50uL 去离子水冲洗两次，14000g，离心 20 分钟。

2.8 蒸干

将收集到的肽段用离心浓缩仪蒸干，并用 0.1% 甲酸水溶液重悬肽段，用于上机分析。

3 仪器条件

3.1 色谱条件

色谱系统为 Thermo Fisher Easy nLC 1200，流动相 A 为 0.1% 甲酸水溶液，B 向为 80% 乙腈 + 20% 水 + 0.1% 甲酸。预柱和分析柱均为实验室自制，外径 360 μ m，内径 150 μ m，预柱长度 2cm，填料粒径为 3 μ m，分析柱长度 30cm，填料粒径为 1.9 μ m。流速为 600nL/min。所用梯度见表 1:

表 1 色谱梯度

时间 min	%B
0	8
10	12
79	27
107	45
110	95
120	95

3.2 质谱参数

质谱系统为 Thermo Fisher Orbitrap Fusion Lumos 三合一质谱仪。经系统考察不同参数的影响，确定质谱参数和 DIA 窗口设置如下，见表 2-4。

表 2 一级质谱参数

名称	参数
Detector Type	Orbitrap
Orbitrap Resolution	60K
Mass Range	Normal
Use Quadrupole Isolation	Yes
Scan Range(m/z)	350-1000
Normalized AGC Target (%)	250

Maximum Injection Time (ms)	50
Data Type	Profile
Polarity	Positive

表 3 二级质谱参数(tMSn 模块)

名称	参数
MSn Level (n)	2
Isolation Mode	Quadrupole
Isolation Window	Defined in Table
Activation Type	HCD
HCD Collision Energy(%)	32
Detector Type	Orbitrap
Orbitrap Resolution	30K
Mass Range	Normal
Scan Range(m/z)	200-2000
Normalized AGC Target (%)	Defined in Table
Maximum Injection Time(ms)	54
Data Type	Centroid
Polarity	Positive

表 4 DIA 窗口设置

Isolation Window (m/z)	AGC Target (Absolute)	m/z	z	t start (min)	t stop (min)
40	500000	370	3	0	120
30	500000	405	3	0	120
30	500000	435	3	0	120
15	500000	457.5	3	0	120
15	500000	472.5	3	0	120
15	500000	487.5	3	0	120
15	500000	502.5	3	0	120
15	500000	517.5	3	0	120
15	500000	532.5	3	0	120
15	500000	547.5	3	0	120
15	500000	562.5	3	0	120
15	500000	577.5	3	0	120

15	500000	592.5	3	0	120
15	500000	607.5	3	0	120
15	500000	622.5	3	0	120
15	500000	637.5	3	0	120
15	500000	652.5	3	0	120
15	500000	667.5	3	0	120
15	500000	682.5	3	0	120
15	500000	697.5	3	0	120
15	500000	712.5	3	0	120
15	500000	727.5	3	0	120
15	500000	742.5	3	0	120
15	500000	757.5	3	0	120
15	500000	772.5	3	0	120
15	500000	787.5	3	0	120
15	500000	802.5	3	0	120
15	500000	817.5	3	0	120
15	500000	832.5	3	0	120
15	500000	847.5	3	0	120
15	500000	862.5	3	0	120
15	500000	877.5	3	0	120
15	500000	892.5	3	0	120
20	500000	910	3	0	120
20	500000	930	3	0	120
20	500000	950	3	0	120
20	500000	970	3	0	120
20	500000	990	3	0	120

4 结果与讨论

本研究采用 HEK293T 细胞来源的酶解肽段，使用纳升液相色谱-高分辨质谱仪对其进行 DIA 模式数据采集。利用优化后的质谱参数，采集不同上样量（500ng、100ng、20ng、4ng）和气相分馏（Gas Phase Fractionation, GPF）^[10]的 DIA 质谱数据，并采集分馏样本的 DDA 数据。利用 DIA-NN^[11]、EncyclopeDIA、Skyline 和 Spectronaut 四种 DIA 数据分析软件，结合三种不同的建库策略，即深度学习谱图库预测、气相分馏谱图库经验校准和 DDA 建立谱图库，共 8 种分析流程对蛋白质定性和非标记定量结果进行系统比较。比较了不同软件 and 不同建库策略的鉴定蛋白数量以及定量的重现性等，以确定最佳分析流程。结果表明，GPF 谱图库经验校准策

略应用于 DIA-NN 软件表现最优，该方案鉴定到的蛋白和肽段数量较多，并且相比于深度学习谱图库预测和 DDA 建立谱图库策略，保留时间校准效果最为明显。

在此基础上，我们对脂多糖（Lipopolysaccharide,LPS）刺激前后的 5000 个外周血单核细胞（Peripheral Blood Mononuclear Cell,PBMC）进行了非标记定量蛋白质组学研究，共检测到 4733 个蛋白，并对其中 3213 个蛋白进行了定量。通过 KEGG 功能富集分析，激活的通路主要包括 T 细胞受体信号通路、NF-kappa B 信号通路、内吞作用等。基因集富集分析（Gene Set Enrichment Analysis,GSEA）结果表明，LPS 刺激后富集到的蛋白最多且激活的通路是免疫系统的细胞因子信号通路。

该研究结果已于 2022 年发表在 Journal of Pharmaceutical and Biomedical Analysis 杂志，题目为 A data-independent acquisition (DIA)-based quantification workflow for proteome analysis of 5000 cells。基于 DIA 的预测谱图库策略可实现微量蛋白质组学的定性定量分析，为微量蛋白质组学的研究提供了新思路，也为临床精准医疗的开展提供了更多可能性。

6 服务成果

多组学质谱实验室成立于 2018 年，拥有 Thermo Fisher 三合一质谱 Orbitrap Fusion Lumos 和 Q Exactive HF 两台超高分辨质谱仪，可用于蛋白组学、代谢组学、脂质组学等研究，并配备了 Bruker ultrafleXtreme MALDI-TOF/TOF 质谱仪，既可用于小分子、生物大分子、合成聚合物的快速分析，也可用于不同组织（包括植物）的质谱成像等，总价值超过 1500 万元。目前已建立了成熟的样品前处理、仪器方法、数据分析的全流程多组学检测分析方法，研究方向包括药物化学结构确证、中药中活性成分研究、药物代谢机理和途径、药物作用靶点和机制研究、药物质量控制研究、精准医学模式下的临床治疗指导方案等。截至目前，已培训学生超 300 人次，支持和合作科研项目 100 余项，协助校内外用户开展了多项高水平科研工作，支撑发表 SCI 文章 67 篇，研究成果见于 *Science* (2021,372:512-516)、*Angew. Chem. Int. Ed.* (2020,59:3994-3999,封面)、*Anal. Chem.* (2023,95:7702-7714) 等。

7 社会效益

2019 年 3 月受应急管理部消防救援局天津火灾物证鉴定中心的委托，质谱实验室对“3.21”响水重大爆炸事故现场物证进行快速准确鉴定，受到委托单位的高度肯定，并送来了感谢信。

本实验室参与的“生物质谱分析新方法开发及其在医药转化中的应用”项目获得了 2021 年度中国分析测试协会科学技术奖二等奖（省部级）。

为持续提升专业水平与管理品质，更好地为科研与教学提供高质量的技术支持与服务，包括多组学质谱实验室在内的仪器测试中心于 2022 年 7 月正式获得中国合格评定国家认可委员会的 CNAS/iLAC-MRA 双重认证，成为 CNAS 认可的检测实验室（注册号：CNASL16629）。

8 总结

多组学质谱实验室深入地交叉整合了分析化学、药学、生物信息学、细胞与分子生物学、临床医学等多个学科的技术手段，以解决精准医疗中所面临的重要问题，建立了具有普适性、易操作、可重复的生物质谱新方法，以用于发现新的药物治疗靶点，提出新的诊疗思路，具有很高的推广价值。

参考文献:

- [1] Clarke NJ. Mass Spectrometry in Precision Medicine: Phenotypic Measurements Alongside Pharmacogenomics. *Clin Chem*, 2016, 62(1):70-6.
- [2] Azad RK, Shulaev V. Metabolomics technology and bioinformatics for precision medicine. *Brief Bioinform*, 2019, 20(6):1957-1971.
- [3] König IR, Fuchs O, Hansen G, et al. What is precision medicine? *Eur Respir J*, 50(4):1700391.
- [4] Gaspar VP, Ibrahim S, Zahedi RP, et al. Utility, promise, and limitations of liquid chromatography-mass spectrometry-based therapeutic drug monitoring in precision medicine. *J Mass Spectrom*, 2021, 56(11):e4788.
- [5] Longuespée R, Theile D, Zörnig I, et al. Molecular prediction of clinical response to anti-PD-1/anti-PD-L1 immune checkpoint inhibitors: New perspectives for precision medicine and mass spectrometry-based investigations. *Int J Cancer*, 2023, 153(2):252-264.
- [6] Bravo-Veyrat S, Hopfgartner G. Mass spectrometry based high-throughput bioanalysis of low molecular weight compounds: are we ready to support personalized medicine? *Anal Bioanal Chem*, 2022, 414(1):181-192.
- [7] Teclemariam ET, Pergande MR, Cologna SM. Considerations for mass spectrometry-based multi-omic analysis of clinical samples. *Expert Rev Proteomics*, 2020, 17(2):99-107.
- [8] Pino LK, Just SC, MacCoss MJ, et al. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol Cell Proteomics*, 2020, 19(7):1088-1103.
- [9] Wiśniewski JR, Zougman A, Nagaraj N, et al. Universal sample preparation method for proteome analysis. *Nat Methods*, 2009, 6(5):359-62.
- [10] Searle BC, Swearingen KE, Barnes CA, et al. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat Commun*, 2020, 11(1):1548.
- [11] Demichev V, Messner C B, Vernardis S I, et al. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods*, 2020, 17(1):41-44.