

面向亲和富集-质谱分析通用需求的生物信息学 分析平台

张骞宁^{1,2}, 何岸^{1,2}, 李元^{1,2}, 柯弥^{1,2}, 田瑞军^{1,2}

(1. 南方科技大学理学院化学系, 广东 深圳 518055; 2. 深圳市功能蛋白质组学重点实验室, 广东 深圳 518055)

摘要: 亲和富集-质谱 (affinity purification-mass spectrometry, AP-MS) 技术是解析蛋白质-蛋白质相互作用的重要手段。随着质谱仪器性能的持续提升, 高质量 AP-MS 数据得以快速积累, 然而, 相应数据分析工具的开发和更新尚未与之同步。为此, 本研究构建了一款多功能数据分析 R 语言包——APMSflow, 其封装了常见的处理蛋白质组学数据的函数, 用户无需重复写代码, 即可高效处理 AP-MS 数据。该工具集成了质量控制、数据预处理、差异分析、下游功能分析、蛋白质互作网络和预测等多个模块, 覆盖了 AP-MS 数据分析的核心流程。APMSflow 设计旨在助力提升对质谱数据的处理能力, 为深入解析蛋白质相互作用及重建蛋白质复合物网络提供便捷、高效的解决方案。

关键词: 亲和富集-质谱 (AP-MS); 蛋白质-蛋白质相互作用; 数据分析

中图分类号: O657.63 文献标志码: A 文章编号: 1004-2997(2026)03-0456-08

DOI: 10.7538/zpxb.2026.0028

CSTR: 32365.14.zpxb.2026.0028

A Bioinformatics Platform for General Affinity Purification-Mass Spectrometry Analysis

ZHANG Qian-ning^{1,2}, HE An^{1,2}, LI Yuan^{1,2}, KE Mi^{1,2}, TIAN Rui-jun^{1,2}

(1. Department of Chemistry, School of Science, Southern University of Science and Technology, Shenzhen 518055, China; 2. Shenzhen Key Laboratory of Functional Proteomics, Shenzhen 518055, China)

Abstract: Protein-protein interactions (PPIs) are fundamental to cellular regulation, governing complex biological processes through the assembly of stable or transient molecular machineries. Affinity purification coupled with mass spectrometry (AP-MS) has emerged as the gold standard for characterizing these interactomes. However, as AP-MS datasets grow in heterogeneity and complexity—incorporating both data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes as well as multi-bait and time-series experimental designs—researchers face significant challenges. These include high levels of non-specific binding, complex background noise, and the lack of standardized, end-to-end computational pipelines that integrate quality control with advanced structural prediction. In this study, APMSflow, a comprehensive, R/Shiny-based analytical platform was developed, specifically engineered for the systematic processing of AP-MS data. The workflow integrates several critical modules: 1) Quality assessment, utilizing PCA and peptide-level metrics to

evaluate sample reproducibility and digestion efficiency; 2) Pre-processing, offering multiple normalization strategies (e.g., median, bait-based, or endogenous biotinylated protein-based) and sophisticated missing value imputation methods; 3) Differential analysis, implementing moderated *t*-tests *via* the Limma package and a novel “multi-bait background modeling” strategy to enhance interaction specificity without requiring independent negative controls; 4) Downstream integration, combining cluster-based functional enrichment (clusterProfiler) with structural bioinformatics analysis. Uniquely, APMSflow incorporates AlphaFold-Multimer to perform structural modeling and scoring of predicted protein pairs, transitioning from statistical association analysis to structural validation. Validation performed using published large-scale proteomics datasets (PXD020709) demonstrated that APMSflow effectively filters non-specific contaminants and identifies high-confidence interactors. The platform successfully captured the temporal dynamics of the EGFR interactome, categorizing proteins into transient or stable interaction clusters based on their abundance profiles across time points. By applying stringent CV-based filtering and specialized handling of bait-specific proteins (addressing “NA” values in control groups), APMSflow recovered biologically relevant early-transient signaling components that are often missed by conventional pipelines. The integration of AlphaFold-Multimer further provided structural evidence for candidate protein complexes, streamlining the prioritization of targets for biochemical validation. APMSflow addresses a critical gap in the proteomics community by providing an accessible, standardized, and robust tool for protein interactome analysis. It lowers the technical barrier for wet-lab researchers while ensuring experimental reproducibility across studies. While current iterations focus on protein-level quantification, future updates aim to incorporate post-translational modification (PTM) site-specific analysis and more advanced protein complex prediction algorithms. APMSflow represents a significant step toward the automated, structure-aware interpretation of the dynamic protein interaction landscape, facilitating the discovery of novel regulatory mechanisms in systems biology. The tool is freely accessible as a web application at: <https://humility3238.shinyapps.io/apmsflow/>.

Key words: affinity purification-mass spectrometry (AP-MS); protein-protein interactions; data analysis

细胞功能受到蛋白质-蛋白质相互作用的严格调节^[1-2]。这些蛋白质复合物包括间接或直接的蛋白质复合物、胞外基质蛋白质复合物、翻译后修饰依赖性蛋白质复合物、时空动态蛋白质复合物、信号通路动态组装蛋白质复合物等。目前,研究这些蛋白质-蛋白质相互作用的主流策略主要包括基于抗体偶联微球的免疫共沉淀富集^[3-5]、活细胞邻近标记方法^[6-8]、以及有机合成肽段富集的蛋白质相互作用方法^[9-10]等。这些蛋白质复合物富集策略与高分辨质谱分析结合,1次实验即可获得与目标蛋白质互相作用的所有蛋白质,称为亲和富集-质谱(AP-MS)技术。随着 AP-MS 数据在实验模式和分析深度上的不断扩展,其数据复杂性和异质性日益增加。因此,亟需针对 AP-MS 数据特点开发具备批量处理能力的高效、稳健且可扩展的计算工具。

目前,针对蛋白质组学数据分析已开发出多种软件平台。2010年, Tyers 和 Gingras 团队开发了 ProHits 平台,该平台集成了质谱数据追踪、AP-MS 数据分析与注释等功能,是一个面向多用户、多项目并行管理的综合系统^[11-12]。2016年, Cox 团队推出了 Perseus 平台,具备定量数据预处理、常用统计分析方法及数据可视化等核心功能,并通过图形用户界面实现了分析流程的直观展示^[13]。近年来,基于 RShiny 框架开发的数据分析流程不断涌现,如 ProVision 是一个基于无标记(label-free)和 TMT (tandem mass tag)的定量分析网络应用程序^[14]; ProteoMill 整合了统计方法和下游的功能注释流程^[15]; LFQ-Analyst 主要针对无标定量数据的差异分析和可视化的需求进行设计^[16]; DEP2 继承了 Perseus 常规分析框架,并进一步集成了缺失值填补、数据归一化及

多样化可视化方法,显著提升了数据处理的灵活性与可视化效果^[17]。尽管相关分析工具不断丰富,但随着 AP-MS 数据量的快速增长及实验设计日益复杂,现有分析流程在适用性和扩展性方面逐渐难以满足实际研究需求。

与常规定量蛋白质组学数据相比, AP-MS 数据具有背景信号复杂、非特异性结合普遍等特点。在实际分析过程中,如何合理构建背景对照、降低非特异性信号干扰,并在多诱饵或复杂实验设计下实现稳健的差异分析,是 AP-MS 数据分析面临的关键挑战。此外,分析流程往往依赖多种分散的软件工具和较多的人工经验,增加了不同研究之间的结果差异性,不利于流程的标准化与结果的可重复性。

为此,本研究开发了一款专为 AP-MS 数据设计的分析平台,即 APMSflow,覆盖从数据质量评估、预处理、差异分析、通路分析、蛋白质互作网络构建及预测的全流程。该平台兼容数据依赖采集(data-dependent acquisition, DDA)与数据独立采集(data-independent acquisition, DIA)格式输入,具备良好的适用性。

1 实验部分

本研究所使用的 R 语言包在 Windows 10 操作系统中开发完成,并基于 R 软件(版本 R 4.4.2)实现。分析数据来源于已公开发布的数据集,存储于 ProteomeXchange 数据库,编号为 PXD020709^[6]。APMSflow 在线访问地址为 <https://humility3238.shinyapps.io/apmsflow/>,其功能界面主要划分为数据上传、参数配置以及结果可视化与下载 3 个阶段。整体分析流程及相应功能列于表 1。用户登录网址后,首先上传蛋白质组

表 1 AP-MS 数据分析流程和实现功能
Table 1 AP-MS data analysis workflow and related functions

模块 Module	主要实现功能 Function
数据质量评估	蛋白质和肽段鉴定量展示,蛋白质重复性评估
数据预处理	过滤反库和污染库蛋白质,表达量校正,缺失值替代
差异表达分析	双样本 t 检验
诱饵特异性蛋白质分析	筛选互作蛋白质
下游功能分析	GO, KEGG 分析
蛋白质相互作用网络构建	蛋白质复合物预测

数据库检索获得的定量文件及实验设计文件;随后设置分析参数,包括阈值设定、数据校正方法、缺失值填补策略以及诱饵蛋白质名称等;完成参数配置后,用户可一键启动分析流程,系统将自动生成相应的数据表格及可视化结果,示于图 1a。

2 结果与讨论

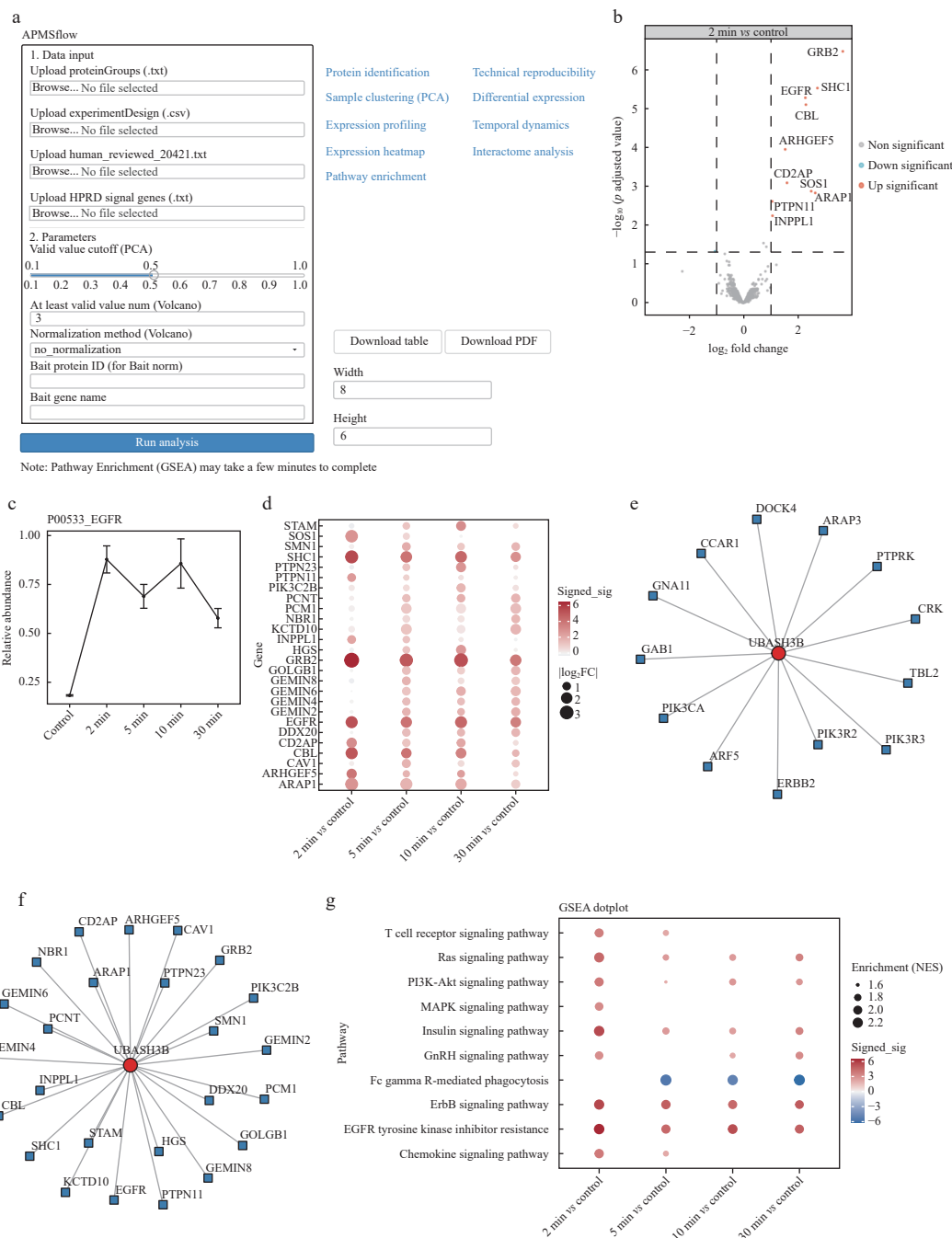
本研究构建了一套面向 AP-MS 数据的系统分析流程,涵盖数据质量评估、数据预处理与统计分析等关键步骤,以支持复杂实验设计下的稳健差异分析。该流程能够适应多诱饵条件的比较需求,并通过合理的背景建模提高特异性互作蛋白质的识别能力,从而为后续分析提供可靠的候选蛋白质集合。在此基础上,进一步引入功能注释与蛋白质相互作用网络构建及预测等分析模块,对筛选得到的候选蛋白质进行多层次综合评估,最终生成具有较高生物学相关性的蛋白质列表,为后续实验验证与机制研究提供参考。

APMSflow 旨在降低实验室研究人员在质谱数据分析方面的技术门槛,提供一套集成化的数据处理流程,以简化对具有生物学意义的蛋白质互作网络的识别与解析。该平台围绕 AP-MS 数据分析需求构建了若干个核心功能模块,包括数据质量控制、数据预处理、差异分析、功能分析和蛋白质相互作用网络构建及预测,使研究者能够高效地将原始质谱数据转化为可解释的生物学结果,示于图 2。

2.1 数据质量评估模块

在数据质量评估模块中,用户首先需输入谱图(spectrum)层面的定性数据,例如 MaxQuant 软件数据库检索结果生成的 evidence.txt 文件^[18]。系统基于该数据自动生成酶切效率及肽段电荷态分布等质量控制报告,用于评估样本制备过程的可靠性;随后,需引入蛋白质层面的定量数据(如 MaxQuant 导出的 proteinGroups.txt 文件),对肽段鉴定数、蛋白质鉴定数以及蛋白质定量变异系数(coefficient of variation, CV)分布进行统计分析,从而识别潜在的离群样本。

此外,通过主成分分析(principal component analysis, PCA)对样本间的整体相似性进行评估,以进一步检验实验重复性的稳定性。蛋白质丰度动态范围分布图用于确认样本中所鉴



注：a. APMSflow 软件图形用户界面 (GUI) 展示；b. 两组差异比较分析生成的火山图；c. 显著差异蛋白 EGFR 丰度变化趋势折线图；d. 各比较组中显著上调蛋白的 \log_2 倍数变化 ($\log_2 \text{ fold change}$) 分布点图；e. 与诱饵蛋白 (Bait) UBASH3B 具有瞬时相互作用的蛋白质组图谱；f. 与诱饵蛋白 UBASH3B 存在早期响应关系的蛋白质组图谱；g. 各比较组蛋白质富集分析 (GSEA) 结果点图

图 1 APMSflow 平台界面及代表性分析结果可视化

Fig. 1 APMSflow interface and representative analysis visualizations

定蛋白质覆盖的从低丰度到高丰度的广泛数量级范围。

2.2 数据预处理模块

在完成初始质量控制流程后, 将进入数据预处理模块。首先, 对蛋白质层面的定量数据进行过滤, 去除反向数据库 (reverse database)、污染数

据库 (potential contaminants) 以及仅在位点层面鉴定到 (only identified by site) 的蛋白质; 随后, 根据实验设计对样本进行分组, 并对定量强度值进行对数转换; 在此基础上, 筛选至少在任一实验组中具有不少于指定有效值数量的蛋白质, 其中各组所需的有效值数量可由用户自行设定; 接

着,引入多种归一化(normalization)策略,以校正不同样本间的技术差异,确保组间数据的可比性,常用方法包括中位数归一化、基于内源性生物素化蛋白质的归一化以及基于诱饵蛋白质的归一化^[6,19];最后,流程整合多种缺失值替代(missing value imputation)方法^[20],用户可根据数据特征与分析目标选择合适的替代策略。

需要注意的是,在分析基于磷酸化肽段的蛋白质相互作用数据时,通常首先需对诱饵蛋白质自身的肽段信号强度进行单独分析,其中的磷酸化肽段在后续诱饵蛋白质整体信号强度计算中予以剔除,随后仅基于剩余肽段重新计算诱饵蛋白质丰度,以避免磷酸化修饰对定量结果的干扰。

2.3 差异分析模块

在差异分析模块中,可采用 Limma 软件包中的 moderated *t*-test 或传统的双样本 ordinary *t*-test 对蛋白质定量数据进行差异分析^[21],结果示于图 1b。当输入数据包含多个诱饵蛋白质样本时,可将其与研究对象无关的诱饵蛋白质样本合并作为整体背景组,并与目标诱饵蛋白质样本进行比较^[22-23]。该策略避免了额外设置独立对照组的需求,同时利用多样本背景信息进行统计建模,有助于提高差异倍数估计与显著性检验结果的稳健性和可靠性。

在获得多组差异分析结果后,首先提取在多次比较中均显著上调的蛋白质;随后,对这些蛋白质在不同时间点的定量结果进行聚类分析,将时间动态变化趋势相似的蛋白质归为同一类,示于图 1b。在此基础上,对每个蛋白质的表达动态进行单独展示,示于图 1c。结果表明,EGFR 蛋白质在 2 min 时达到表达峰值,在 5~10 min 期间呈轻微上升趋势,随后于 30 min 回落。该类蛋白质的时间动态曲线可作为直观参考反馈给实验研究人员,用于指导后续的功能验证与机制研究。将多组比较获得的差异蛋白质进行合并,并绘制各比较组相对于对照组变化倍数的点图,示于图 1d。该点图能够清晰展示蛋白质在不同比较组中的变化趋势,有助于发现变化趋势相似的蛋白簇。

在 AP-MS 数据分析中,受质谱检测限或蛋白质丰度较低等因素影响,部分蛋白质在对照组的多次重复样本中可能未被检测到。为避免潜在在相互作用蛋白质的遗漏,本研究针对诱饵特异

性蛋白质建立了一套完整的分析流程。在仅涉及 2 组比较的情况下,对该类蛋白质进行单独筛选,并结合其在实验组中的变异系数进行排序评估,以决定是否保留进入后续的分析流程。若该蛋白质在实验组中的变异系数值较低且排序靠前,表明其定量结果具有良好的重复性,则予以保留。将本步骤筛选得到的蛋白质与每次差异比较中显著上调的蛋白质分别整合,构建用于后续功能富集分析的蛋白质集合。在多组实验数据比较中,首先分别对各时间点进行统计检验,并与对照组逐一比较,获得显著上调蛋白及其他蛋白质。对于非显著上调蛋白,进一步筛选对照组全部为 NA(not available)且在第二时间点达到均值峰值的蛋白质,并与 HPRD (The Human Protein Reference Database)和 Uniprot 数据库中具有 Signal Transduction 注释的蛋白质取交集,得到早期瞬时互作蛋白质组,示于图 1e。对于显著上调蛋白,通过严格的 CV 阈值筛选,并要求至少连续 4 个时间点具有可计算的 CV 值,从而构建时间动态互作蛋白质组,示于图 1f。通过上述流程,最终获得高置信度的蛋白质列表。

2.4 下游功能分析模块

功能注释与通路分析可基于成熟的 clusterProfiler R 软件包完成^[24],示于图 1g。该工具不仅适用于单一条件下的富集分析,在涉及多个时间点或多个诱饵蛋白质的数据比较时亦具有良好的适用性。分析过程中,需重点关注随时间进程或诱饵蛋白质的变化及各功能通路富集程度的动态变化趋势,包括增强、稳定或减弱等不同模式。通过对通路富集结果进行可视化整合,可在同一张图中展示通路层面的动态变化,从而有助于揭示潜在的生物学调控过程及其内在机制。

2.5 蛋白质相互作用网络构建及预测模块

在蛋白质相互作用网络构建与预测模块中,首先将差异显著的蛋白质整合构建为蛋白质相互作用网络,并基于功能相似性对相关蛋白质进行聚类分析,以识别潜在的功能模块。随后,将各诱饵蛋白质与其候选互作蛋白质组成的蛋白质对导入 AlphaFold-Multimer^[25-26],对潜在蛋白质相互作用进行结构建模与打分评估。最终,所有网络分析、结构预测及功能注释结果将以可视化报告形式呈现给实验研究人员,为其筛选具有

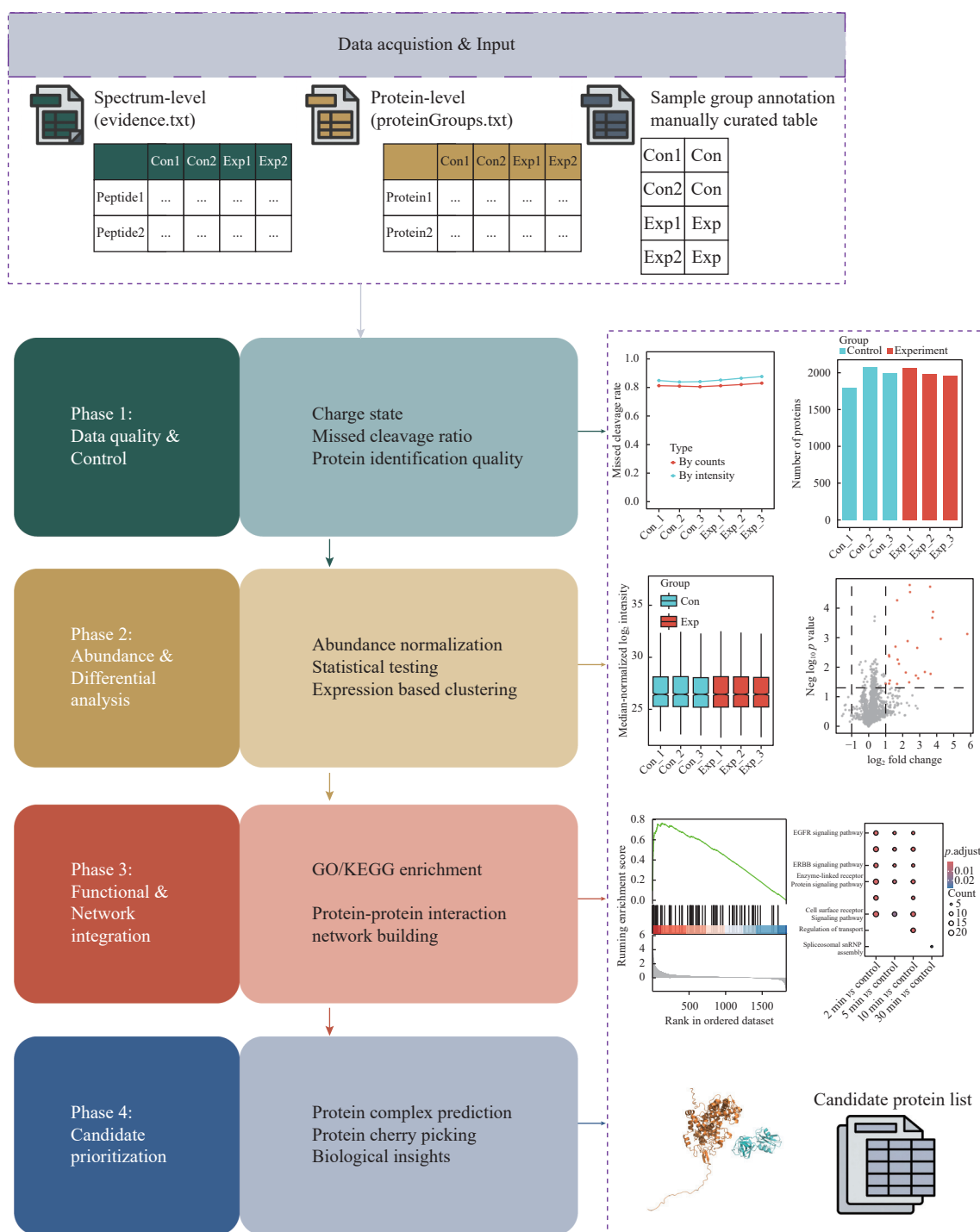


图 2 基于 AP-MS 蛋白质组学数据类型的 APMSflow 分析流程示意图

Fig. 2 Schematic workflow of APMSflow analysis based on AP-MS proteomics data types

较高生物学相关性的候选蛋白质,以开展进一步实验验证提供参考依据。

3 结论

本研究开发了一款面向 AP-MS 数据处理与蛋白质相互作用网络构建的多功能计算分析平

台——APMSflow。该平台通过系统整合差异分析与蛋白质相互作用网络构建流程,有效提升了 AP-MS 数据分析的可重复性与生物学可解释性,使研究人员能够在较短时间内获得初步的蛋白质相互作用网络,从而加快候选互作蛋白质的筛选过程,并为后续机制研究与功能验证实验提

供有力支持。

尽管 APMSflow 在简化 AP-MS 数据分析流程方面表现出良好的实用性与通用性,但当前版本仍存在一定局限性,如尚未覆盖所有归一化策略与差异分析方法,且暂不支持翻译后修饰(如磷酸化)位点层面的数据分析。未来的版本迭代将重点引入更加完善的蛋白质复合物预测方法,并进一步整合翻译后修饰肽段层面的信息,以增强平台在大规模蛋白质相互作用组学研究中的适用性与扩展能力。

参考文献:

- [1] HENTZE M W, CASTELLO A, SCHWARZL T, PREISS T. A brave new world of RNA-binding proteins[J]. *Nature Reviews Molecular Cell Biology*, 2018, 19(5): 327-341.
- [2] KESKIN O, TUNCBAG N, GURSOY A. Predicting protein-protein interactions from the molecular to the proteome level[J]. *Chemical Reviews*, 2016, 116(8): 4 884-4 909.
- [3] ZHENG Y, ZHANG C, CROUCHER D R, SOLIMAN M A, ST-DENIS N, PASCULESCU A, TAYLOR L, TATE S A, HARDY W R, COLWILL K, DAI A Y, BAGSHAW R, DENNIS J W, GINGRAS A C, DALY R J, PAWSON T. Temporal regulation of EGF signalling networks by the scaffold protein Shc1[J]. *Nature*, 2013, 499(7 457): 166-171.
- [4] BISSON N, JAMES D A, IVOSEV G, TATE S A, BONNER R, TAYLOR L, PAWSON T. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor[J]. *Nature Biotechnology*, 2011, 29(7): 653-658.
- [5] SWANEY D L, RAMMS D J, WANG Z, PARK J, GOTO Y, SOUCHERAY M, BHOLA N, KIM K, ZHENG F, ZENG Y, MCGREGOR M, HERRINGTON K A, O'KEEFE R, JIN N, VANLANDINGHAM N K, FOUSSARD H, VON DOLLEN J, BOUHADDOU M, JIMENEZ-MORALES D, OBERNIER K, KREISBERG J F, KIM M, JOHNSON D E, JURA N, GRANDIS J R, GUTKIND J S, IDEKER T, KROGAN N J. A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity[J]. *Science*, 2021, 374(6 563): eabf2911.
- [6] KE M, YUAN X, HE A, YU P, CHEN W, SHI Y, HUNTER T, ZOU P, TIAN R. Spatiotemporal profiling of cytosolic signaling complexes in living cells by selective proximity proteomics[J]. *Nature Communications*, 2021, 12: 71.
- [7] RHEE H W, ZOU P, UDESHI N D, MARTELL J D, MOOTHA V K, CARR S A, TING A Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging[J]. *Science*, 2013, 339(6 125): 1 328-1 331.
- [8] ROUX K J, KIM D I, RAIDA M, BURKE B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells[J]. *The Journal of Cell Biology*, 2012, 196(6): 801-810.
- [9] CHEN X, JI S, LIU Z, YUAN X, XU C, QI R, HE A, ZHAO H, SONG H, XIAO C, GAO W, CHEN P, LUO R, LI P, WANG F, YANG X, TIAN R. Motif-dependent immune co-receptor interactome profiling by photoaffinity chemical proteomics[J]. *Cell Chemical Biology*, 2022, 29(6): 1 024-1 036. e5.
- [10] CHEN P, CHEN X, SONG X, HE A, ZHENG Y, LI X, TIAN R. Dissecting phospho-motif-dependent Shc1 interactome using long synthetic protein fragments[J]. *Chemical Science*, 2024, 15(36): 14 806-14 815.
- [11] KNIGHT J D R, CHOI H, GUPTA G D, PELLETIER L, RAUGHT B, NESVIZHSHKII A I, GINGRAS A C. ProHits-viz: a suite of web tools for visualizing interaction proteomics data[J]. *Nature Methods*, 2017, 14(7): 645-646.
- [12] LIU G, ZHANG J, LARSEN B, STARK C, BREITKREUTZ A, LIN Z Y, BREITKREUTZ B J, DING Y, COLWILL K, PASCULESCU A, PAWSON T, WRANA J L, NESVIZHSHKII A I, RAUGHT B, TYERS M, GINGRAS A C. ProHits: integrated software for mass spectrometry-based interaction proteomics[J]. *Nature Biotechnology*, 2010, 28(10): 1 015-1 017.
- [13] TYANOVA S, TEMU T, SINITYN P, CARLSON A, HEIN M Y, GEIGER T, MANN M, COX J. The perseus computational platform for comprehensive analysis of (pro) omics data[J]. *Nature Methods*, 2016, 13(9): 731-740.
- [14] GALLANT J L, HEUNIS T, SAMPSON S L, BITTER W. ProVision: a web-based platform for rapid analysis of proteomics data processed by MaxQuant[J]. *Bioinformatics*, 2020, 36(19): 4 965-4 967.
- [15] RYDÉN M, ENGLUND M, ALI N. ProteoMill: efficient network-based functional analysis portal for proteomics data[J]. *Bioinformatics*, 2021, 37(20): 3 491-3 493.
- [16] SHAH A D, GOODE R J A, HUANG C, POWELL D R, SCHITTENHELM R B. LFQ-analyst: an easy-to-use

- interactive web platform to analyze and visualize label-free proteomics data preprocessed with MaxQuant[J]. *Journal of Proteome Research*, 2020, 19(1): 204-211.
- [17] FENG Z, FANG P, ZHENG H, ZHANG X. DEP2: an upgraded comprehensive analysis toolkit for quantitative proteomics data[J]. *Bioinformatics*, 2023, 39(8): btad526.
- [18] COX J, MANN M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification[J]. *Nature Biotechnology*, 2008, 26(12): 1367-1372.
- [19] ZHENG J, ZHENG Z, FU C, WENG Y, HE A, YE X, GAO W, TIAN R. Deciphering intercellular signaling complexes by interaction-guided chemical proteomics[J]. *Nature Communications*, 2023, 14: 4138.
- [20] WANG S, LI W, HU L, CHENG J, YANG H, LIU Y. NAGuideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses[J]. *Nucleic Acids Research*, 2020, 48(14): e83.
- [21] RITCHIE M E, Phipson B, Wu D, Hu Y, Law C W, Shi W, Smyth G K. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Research*, 2015, 43(7): e47.
- [22] LUNDBY A, FRANCIOSA G, EMDAL K B, REFSGAARD J C, GNOSA S P, BEKKER-JENSEN D B, SECHER A, MAURYA S R, PAUL I, MENDEZ B L, KELSTRUP C D, FRANCAVILLA C, KVEIBORG M, MONTOYA G, JENSEN L J, OLSEN J V. Oncogenic mutations rewire signaling pathways by switching protein recruitment to phosphotyrosine sites[J]. *Cell*, 2019, 179(2): 543-560. e26.
- [23] HEIN M Y, PENG D, TODOROVA V, McCARTHY F, KIM K, LIU C, SAVY L, JANUEL C, BALTAZAR-NUNEZ R, SEKHAR M, VAID S, BAX S, VANGIPURAM M, BURGESS J, NJOYA L, WANG E, IVANOV I E, BYRUM J R, PRADEEP S, GONZALEZ C G, ANISEIA Y, CREERY J S, McMORROW A H, SUNSHINE S, YEUNG-LEVY S, DeFELICE B C, MEHTA S B, ITZHAK D N, ELIAS J E, LEONETTI M D. Global organelle profiling reveals subcellular localization and remodeling at proteome scale[J]. *Cell*, 2025, 188(4): 1137-1155. e20.
- [24] YU G, WANG L, HAN Y, HE Q. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *Omics*, 2012, 16(5): 284-287.
- [25] LIU J, GUO Z, WU T, ROY R S, QUADIR F, CHEN C, CHENG J. Enhancing alphaFold-multimer-based protein complex structure prediction with MULTICOM in CASP15[J]. *Communications Biology*, 2023, 6: 1140.
- [26] MIRDITA M, SCHÜTZE K, MORIWAKI Y, HEO L, OVCHINNIKOV S, STEINEGGER M. ColabFold: making protein folding accessible to all[J]. *Nature Methods*, 2022, 19(6): 679-682.
- (收稿日期: 2026-03-29; 修回日期: 2026-04-14)