



评述

基于质谱的定量蛋白质组学策略和方法研究进展

常乘^{①②}, 朱云平^{①②*}

① 军事医学科学院放射与辐射医学研究所, 蛋白质药物国家工程研究中心, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 北京 100850;

② 国家蛋白质科学中心(北京), 北京 102206

* 联系人, E-mail: zhuyunping@gmail.com

收稿日期: 2015-01-29; 接受日期: 2015-03-05

国家重点基础研究发展计划(批准号: 2011CB910601, 2013CB910800)、国际合作计划(批准号: 2014DFB30010)、国家高技术研究发展计划(批准号: 2012AA020409, 2012AA020201)和国家自然科学基金(批准号: 21475150, 21105121)资助项目

doi: 10.1360/N052015-00029

摘要 定量蛋白质组学已经成为组学领域研究的热点之一. 相关实验技术和计算方法不断创新极大地促进了定量蛋白质组学的飞速发展. 常用的定量蛋白质组学策略按照是否需要稳定同位素标记可以分为无标定量和有标定量两大类. 每类策略又产生了众多定量方法和工具, 它们一方面推动了定量蛋白质组学的深入发展; 另一方面, 也在实验策略与技术的发展过程中不断更新. 因此对这些定量实验策略和方法进行系统总结和归纳将有助于定量蛋白质组学的研究. 本文主要从方法学角度全面归纳了目前定量蛋白质组学研究的相关策略和算法, 详述了无标定量和有标定量的具体算法流程并比较了各自特点, 还对以研究蛋白质绝对丰度为目标的绝对定量算法进行了总结, 列举了常用的定量软件和工具, 最后概述了定量结果的质量控制方法, 对定量蛋白质组学方法发展的前景进行了展望.

关键词

质谱

定量蛋白质组学

定量算法

无标定量

有标定量

绝对定量

随着 20 世纪 90 年代生物质谱技术快速发展带来的高通量、高灵敏度和大动态范围等技术优势^[1], 质谱技术已逐渐成为蛋白质组学研究中的主流技术^[2]. 与此同时, 蛋白质组学研究的不断深入使得研究重点也从早期的定性研究逐渐转向定量研究. 目前, 定量蛋白质组学已成为蛋白质组学研究的热点之一. 定量蛋白质组学主要是针对不同状态下蛋白质的表达量进行分析比较, 不仅能鉴定出不同状态下表达的蛋白质, 而且能对其丰度进行精确定量. 这对生物标志物的发现和验证, 以及疾病的诊断、监测和治疗等众多领域都具有重要意义^[3].

近年来, 定量蛋白质组学得到了越来越多的重

视, 可检索的相关文献数目增长迅速(图 1). 不论是实验技术还是相关理论方法, 定量蛋白质组学都取得了飞速的发展和进步. 一方面, 生命组学的研究需求促使定量蛋白质组学的实验策略和技术不断推陈出新, 高分辨率、高精度的生物质谱仪器不断涌现, 为定量蛋白质组学的研究带来了翻天覆地的变化, 不仅能够在定性上解析更多的谱图, 同时也能在定量方面更加深入精确地挖掘质谱数据中的丰富信息; 另一方面, 为满足不同定量实验策略的分析需求, 定量蛋白质组学领域已经产生了众多的定量算法与软件. 这些技术和方法上的进步为定量蛋白质组学的深入研究提供了有力的技术支撑. 目前根据是否对

引用格式: 常乘, 朱云平. 基于质谱的定量蛋白质组学策略和方法研究进展. 中国科学: 生命科学, 2015, 45: 425–438

Chang C, Zhu Y P. Strategies and algorithms for quantitative proteomics based on mass spectrometry. SCIENTIA SINICA Vitae, 2015, 45: 425–438, doi: 10.1360/N052015-00029

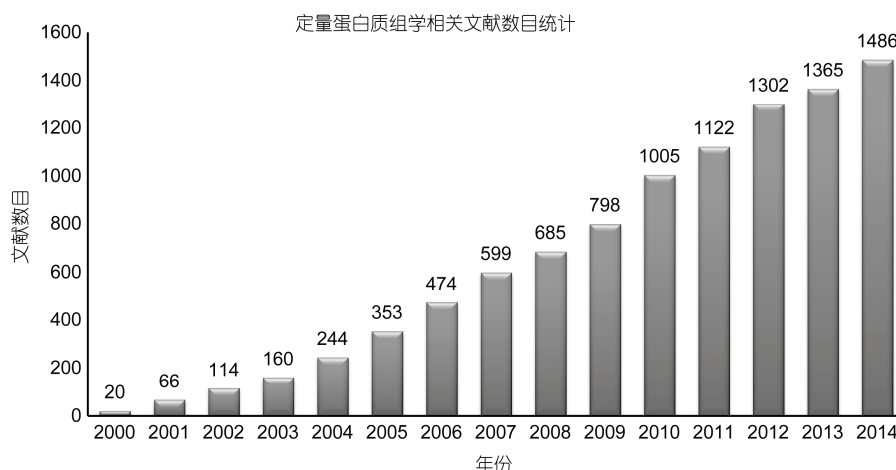


图1 定量蛋白质组学相关文献数目统计

在 pubmed 上搜索关键词“quantitative proteomics”或“quantitative proteome”或“quantification proteome”, 统计 2000 年 1 月 1 日~2014 年 12 月 31 日的相关文献总数

样本进行稳定同位素标记, 定量实验策略和计算方法可以分为无标定量和有标定量两大类. 本文将从方法学的角度详细阐述这两大类方法的原理和具体算法流程并比较了各自特点, 同时评述了以研究蛋白质绝对丰度为目标的绝对定量算法的研究现状, 最后对常用定量软件和工具以及定量结果的质量控制方法进行归纳, 力图对定量蛋白质组学研究中涉及的生物信息学方法进行全面总结, 为相关研究提供参考.

1 无标定量

无标定量(label-free quantification)的计算方法根据不同的定量原理可以分为两类(图 2). 第 1 类是基于离子流色谱峰(extracted ion current, XIC)的定量算法: 在保留时间(retention time, RT)轴上, 根据肽段母离子的质荷比提取不同保留时间下的相应同位素峰簇的信号强度, 重构 XIC, 并利用 XIC 的面积或信号加和等指标作为肽段的定量结果. 要得到 XIC 所对应的肽段序列又有两种方法: (i) 和精确时间质量标签(accurate masses and time tag, AMT)数据库进行匹配^[4]; (ii) 利用蛋白质序列数据库的搜库结果, 确定相应的肽段序列. 第 2 类无标定量方法是谱图计数法(spectral counting, SC)^[5]. 根据蛋白质丰度越高, 对应的肽段被质谱检测到的概率越大的原理, 该方法将一个蛋白质对应的鉴定谱图数(peptide spectrum

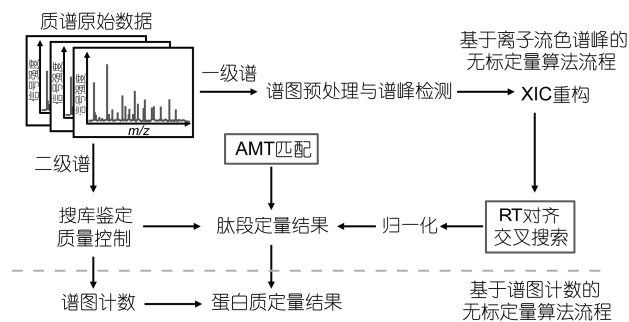


图2 无标定量算法流程图

虚线上方和下方分别是基于 XIC 和基于谱图计数的无标定量算法流程. 其中带框标记的步骤是无标定量算法中的特有步骤

match, PSM)定义为蛋白质的相对丰度. 本文下面将详细介绍这两类无标定量方法的特点与区别.

1.1 基于离子流色谱峰的方法

如图 2 所示, 基于离子流色谱峰的无标定量算法主要包括以下几个主要步骤:

(1) 谱图预处理、谱峰检测与 XIC 重构. 质谱数据主要包含质荷比(m/z)、保留时间(RT)、同位素峰信号强度(isotope intensity) 3 个维度的信息, 可以看成是由多个质荷比、保留时间、同位素峰信号强度组成的三元数据对构成的. 样品经液相分离后在离子源处被离子化而连续地进入质谱(保留时间维度), 质谱仪对离子化的待测离子进行采样和碎裂, 得到相应的一级谱图和二级谱图(质荷比维度和同位素峰信号强度维度). 因此, 谱峰检测和 XIC 重构实际是提

取质谱数据不同维度的信息. 在此之前, 一般要对谱图预处理, 主要是指在定量分析前对质谱数据进行去噪、去基线等操作. 如图 3 所示, 谱峰检测是指通过直接扫描质谱数据中的一级谱图, 在质荷比与信号强度两个维度上, 提取出固定质荷比对应的全部同位素峰各自的信号强度. 该质荷比对应的全部同位素峰称为特征峰型(feature pattern). XIC 重构则是在保留时间和信号强度这两个维度上, 构建出某一质荷比的离子流色谱峰, 具体来说, 就是提取相应质荷比在不同保留时间点上的特征峰型, 该质荷比在不同保留时间上的同位素峰信号强度即为该质荷比对应的 XIC. 张伟等人^[6]对其中的相关算法已经做了详细的综述.

(2) 保留时间对齐与交叉搜索. 在无标定量中, 由于不同状态的样本是分别在质谱中碎裂采样, 因此受限于质谱平台以及实验操作的稳定性, 同一个肽段在不同实验间难免会存在保留时间的偏差. 为了消除这种偏差, 在重构 XIC 之后, 需要对齐保留时间才能在不同实验间互相比对. 保留时间对齐(RT alignment)是无标定量特有的计算步骤之一. 目前常用的对齐算法可以分为以下两类: (i) 在搜库鉴定前基于特征峰型的对齐; (ii) 在搜库鉴定后, 基于鉴定结果的对齐.

早期, 采用线性模型刻画不同样本中的特征峰型/肽段的保留时间对应关系. 但是, 由于液相质谱中肽段的保留时间存在非线性漂移, 因此线性模型会产生较大误差^[7]. 目前, 大多采用非线性模型求解保留时间对齐问题. 例如, MaxLFQ^[8]中提出首先对全部馏分进行层次聚类, 然后对距离最近的馏分两

两对齐, 对齐时采用二维高斯核函数平滑精确时间质量标签之间的保留时间偏移. SuperHirn^[9]则是引用 Silva 等人^[10]的 AMT 对齐算法, 采用聚类的方法将两个不同样本中的特征聚在一起, 然后将这些保留时间漂移进行归一化和 LOWESS 平滑拟合, 得到一个非线性的拟合模型, 用于预测那些未鉴定特征(即未匹配上肽段序列的特征)的保留时间. OpenMS/TOPP^[11]和 MaxLFQ 类似, 是采用了双高斯分布拟合 AMT 的保留时间漂移. Benjamin 等人^[12]采用狄利克雷过程高斯混合模型构建保留时间对齐模型. 该方法和 OpenMS 在 Waters 公司的 MSE 数据上进行了比较, 结果发现狄利克雷过程高斯混合模型更加准确. 除了非线性模型外, 还有使用分而治之算法、半监督式学习等其他方法研究 RT 对齐的工作^[13,14].

基于鉴定结果的保留时间对齐算法是指在搜库鉴定后, 利用鉴定结果进行保留时间的对齐. 其主要优势在于引入了二级谱图的鉴定证据, 结果中的假阳性更少. 同时, 该类方法可以利用肽段序列预测保留时间, 辅助后续交叉搜索. 因此一些无标定量算法, 如 MaxLFQ 实际上是将两类方法结合起来, 虽然基于特征峰型进行保留时间的对齐, 但是只用有搜库鉴定结果的特征峰型作为最终的定量结果.

交叉搜索(cross search)算法也是无标定量中的特有步骤之一. 由于质谱采样的随机性, 在无标定量中, 不同样本会产生不同的鉴定结果, 而保留时间对齐只是针对不同样本中共定量的肽段, 因此对于那些在各自样本中单独定量的肽段, 无法比较其丰度变化. 对于这种类型的肽段, 交叉搜索算法可以利用构建好的保留时间模型预测它们在其他样本中的保留时间, 然后根据该肽段的质荷比提取它们在其他样本中的特征峰型, 重构相应的 XIC, 最后便可计算出该肽段在不同样本中的定量结果. 此算法尽可能地弥补了不同样本间鉴定结果的差异, 提高了数据的可定量比例, 其中的代表性工作是 IDEAL-Q^[15]和 LFQuant^[16].

保留时间对齐和交叉搜索联用, 使得无标定量不再受限于质谱的随机采样, 大大提高了数据的可定量比例. 这里要注意的是, 并不是所有无标定量数据都能进行保留时间对齐与交叉搜索. 一般来说, 其前提是样本间的差异不大或者重复性较好. 若样本间的差异较大或者保留时间的漂移过大, 则可能导致保留时间对齐模型构建失败或交叉搜索时提取相

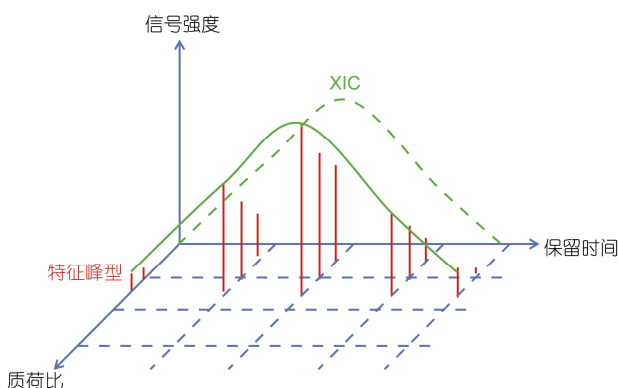


图 3 质谱数据的三维结构

一组同位素峰簇(红色实线)称为一个特征峰型. 不同保留时间上的特征峰型构成了该质荷比对应的 XIC(绿色虚线)

应的特征峰型失败, 这时只能对样本间共定量的肽段进行后续分析。

(3) 归一化. 归一化方法是为了消除不同实验批次间、不同蛋白质之间的系统偏差. 这也是无标定量的必备步骤之一. 目前常用的归一化方法有以下几种^[17]: (i) 全局归一化; (ii) 中值归一化; (iii) 基于内标的归一化; (iv) 分位数归一化; (v) 线性归一化; (vi) 基于稳定特征峰型的归一化, 以每个馏分中变化很小的特征峰型为参照进行归一化。

关于归一化方法评估的工作有不少, 例如, Callister 等人^[18]早在 2006 年就已对常用的归一化方法进行了分析比较, 他们认为肽段的丰度不仅包含实际生物量的变化, 还包含了来自于实验操作和样本制备的误差, 以及仪器的系统误差和随机误差. 经过一系列评估, 最终得出线性归一化方法最优的结论. Webb-Robertson 等人^[19]对不同的归一化方法也进行了比较评估, 试图为每个数据集寻找最合适的归一化方法, 推出了相应的评估软件 SPANS。

(4) XIC 与肽段序列匹配. 在计算肽段的定量结果之前, 首先需要将 XIC 定量信息与相应的肽段序列对应起来. 常用的匹配方法有两种: (i) 根据 AMT 匹配肽段的序列; (ii) 通过搜索蛋白质序列数据库匹配肽段序列. 其中第 2 种方法就是熟知的搜库鉴定问题, 下面重点介绍第 1 种基于 AMT 进行匹配的方法。

如果要在缺少鉴定结果的情况下对齐这些特征峰型, 只能依靠质荷比、保留时间和特征峰信号强度的精确匹配, 一般称此三要素为精确时间质量标签. 该方法适用于高精度、高分辨率的质谱数据(每个特征峰的质荷比一般要精确到 1×10^{-6}). 该方法不局限于搜库鉴定结果, 能够在一定程度上避免质谱随机采样带来的偏差, 得到更多的定量信息. 但是由于在谱峰检测时无法利用肽段序列预测理论同位素峰簇的分布, 该方法有可能引入更多的噪声, 因此在匹配准确性上不如第 2 种基于搜库鉴定的方法。

(5) 肽段与蛋白质定量结果计算. 在基于鸟枪法的实验策略中, 蛋白质的定量信息来源于其对应肽段的定量结果. 目前常用的肽段定量指标计算方法有 XIC 面积法、信号加和法、最大值法、峰型拟合法等^[20]. 需要注意的是使用这些计算方法时要考虑噪声、系统误差对肽段定量结果的影响, 因此一般

先要对 XIC 去噪、平滑并做系统校正, 去除不准确的定量信息。

而蛋白质的定量指标一般有两种, (i) 蛋白质丰度(protein abundance). 在基于离子流色谱峰的方法中, 按照同一个蛋白质对应的肽段权重是否一样, 蛋白质丰度的计算方法可以分为两类: 假设同一个蛋白质对应的全部肽段权重一样, 则蛋白质丰度值的计算方法有对应全部肽段定量值之和、对应全部肽段定量值的平均值和定量值最高的前 3 个肽段的平均值(简称 Top3 法)^[21]等; 假设同一个蛋白对应的肽段权重不同, 在实际分析中, 发现肽段定量值越大, 误差往往越小, 因此可以按照肽段定量值的大小设置相应的权重. 具体计算时, 针对那些包含较多肽段的蛋白质, 还可以选择在 2~3 倍标准差变化范围以内的肽段定量结果, 从而有效地减少蛋白质层面的定量误差; (ii) 蛋白质在不同状态下的相对变化比值(protein ratio). 在比较蛋白质组学或者差异蛋白质组学中, 人们往往关心同一个蛋白质在不同状态之间的差异变化, 因此蛋白质比值的计算直接关系着后续的生物分析. 目前蛋白质比值计算主要分为两种方法: 由蛋白质丰度直接作比得到蛋白质比值, 该方法的前提是同一个蛋白质在不同状态中可定量的肽段是一样的. 在有标定量中, 不同状态的同位素峰簇都在实验时混合在了同一个谱图中, 因此用这类方法可以进行准确定量. 但是在无标定量中, 由于不同实验批次间的蛋白质鉴定结果不同, 因此同一个蛋白质在不同状态中有可能对应不同的肽段. 虽然通过交叉搜索能够部分弥补这种差异, 但是这种情况仍然普遍存在, 所以由蛋白质丰度直接作比的方法并不适用于无标定量. 由肽段丰度比值计算蛋白质的丰度比值, 如取对应肽段比值的中值或者加权平均值. SILVER^[22]中就是采用了 One-step Tukey's Biweight 算法做加权平均, 每个肽段的权重定义该肽段丰度比值与全部肽段丰度比值的中值的距离. 肽段丰度比值距离中值越远, 越有可能是孤点, 权重就越小. 该方法能够克服个别定量不准确的肽段对蛋白质定量结果的干扰, 是目前的常用方法。

此外, 由于生物体进化过程中出现的基因倍增、可变剪接等事件, 部分基因间存在某种程度的相似性, 导致往往会出现一个肽段对应多个蛋白质的情况, 这些肽段称为共享肽段(shared peptide). 由于目

前没有办法在质谱鉴定中区分这些肽段属于哪个蛋白质, 所以通常的做法是忽略共享肽段, 但是这样会丢失部分定量信息, 影响定量准确性. 因此共享肽段的定量信息分配方法直接影响最终的蛋白质推断 (protein inference)^[23]与蛋白质定量结果, 显得十分重要. Zhang 等人^[24]主要研究了共享肽段如何分配更加灵敏和准确的问题, 通过比较 NSAF 算法^[25]的 12 种不同形式的计算结果, 认为按照唯一鉴定肽段 (unique peptide) 的谱图计数值对共享肽段分配能够得到最准确和稳定的效果. Nesvizhskii 和 Aebersold^[26]采用最大期望算法(expectation maximization algorithm)分配共享肽段, 其核心思想是马太效应——对于丰度高的蛋白质, 共享肽段分配比例就高, 低丰度蛋白质分配的比例就低. 但是数学上的最优解不等于生物学的最优解, 例如, 共享肽段的产生可能是样本发生了生物学变化, 也可能是由于鸟枪法的局限性. 因此目前仍然缺少合适的共享肽段分配方法.

蛋白质丰度除了用于定量研究外, 还可以反过来影响定性分析结果, 例如, Lukasse 和 America^[27]创新性地利用肽段在不同状态间的定量变化趋势 (quantification pattern) 区分可能存在的蛋白质亚型, 对通用的简约原则装配出的蛋白质集合 (protein group) 进行拆分, 从而影响最终的蛋白质推断结果.

1.2 谱图计数法

根据质谱采样原理, 一个蛋白质的丰度越高, 酶切出的肽段丰度就越高, 被质谱检测到的概率也越大, 因此将某个蛋白质对应的鉴定谱图数作为表征该蛋白质丰度的指标, 提出了谱图计数法. 因其原理简单、直观, 计算简便, 谱图计数法成为最早被研究者广泛使用的无标定量方法. 在此基础上, 采用不同策略对谱图计数法进行归一化和变换, 提出了 NSAF, APEX^[28], emPAI^[29], SI_N^[30]等优化之后的谱图计数指标, 进一步扩大了谱图计数法在无标定量中的应用范围. 其中的代表性算法是 APEX, 其利用机器学习预测肽段的可检测概率, 用于校正每个蛋白质的谱图数. APEX 与免疫印迹、2D 胶分离以及 mRNA 的丰度计算值一致, 动态范围在 3~4 个数量级, 相比原有的谱图计数法而言, 此算法更加准确.

但是, 由于实验操作和质谱分析等过程中可能存在不可避免的误差, 因此谱图计数法只能称作“半定量”^[31], 这是该方法的最大缺点. 从定量准确性来

看, 谱图计数法不如基于离子流色谱峰的方法. 因此, 也有研究考虑将谱图计数和基于离子流色谱峰的方法相结合. 例如, Wu 等人^[32]将 APEX 中的定量指标由谱图计数改为了蛋白质对应的全部二级谱图的同位素峰信号强度之和, 在癌症临床蛋白质组技术组织 (the Clinical Proteomics Technology Assessment for Cancer, CPTAC) 的标准数据集^[33]上测试的结果表明其效果和 APEX 类似.

2 有标定量

有标定量 (stable isotopic labeling quantification) 是指利用稳定同位素标记样本后再进行相应的质谱鉴定与定量分析的方法. 其原理在于相同肽段的不同标记状态会在同一张质谱图中形成具有固定质量差的同位素峰, 利用这些同位素峰可以计算出同一肽段在每种标记状态中的丰度信息与相应的丰度比值.

按照标记对象的不同, 可以将有标定量分为母离子标记和子离子标记两大类, 其中, 前者是针对肽段母离子进行标记, 在质谱的一级谱图中引入质量差. 后者是针对肽段子离子进行标记, 在质谱二级谱图中对肽段碎片离子引入质量差. 这两种标记方法的定量算法流程如图 4 所示.

2.1 母离子标记定量

母离子标记是目前有标定量中发展时间最早,

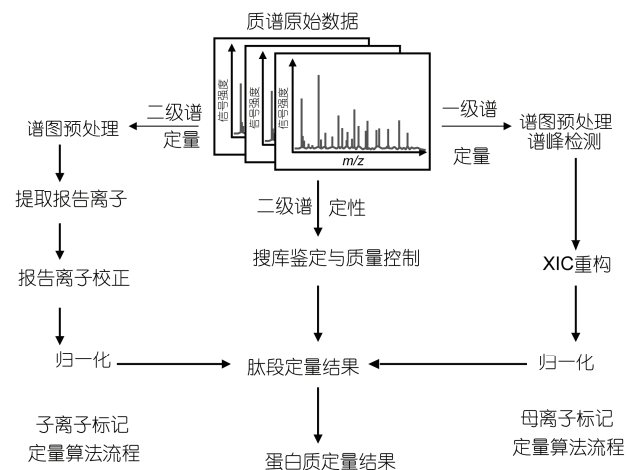


图 4 有标定量算法流程图

左侧是子离子标记的定量算法流程, 右侧是母离子标记的定量算法流程

种类最多的一类标记方法。具体来说,母离子标记又主要分为代谢标记和化学标记两类。代谢标记是让样品在富含某种稳定同位素或由稳定同位素标记的必需氨基酸的培养基中进行培养,经过若干次细胞倍增后,产出蛋白质的元素或氨基酸将由培养基中所富含的稳定同位素或必需氨基酸代替,这样便引入了相应的标记。常见的代谢标记方法有基于元素标记的 ^{15}N 标记法和 ^{13}C 标记法,基于氨基酸标记的SILAC(stable isotope labeling by amino acids in cell culture)^[34]标记法。其中, ^{15}N ^[35,36]和 ^{13}C 标记与其他母离子标记方法在定量计算时最大的不同点在于, ^{15}N 和 ^{13}C 标记的肽段质量差异由肽段所含的N或C元素个数决定,因此不同标记肽段的质量差也会不同。SILAC标记是在细胞培养时加入稳定同位素标记的必需氨基酸,现在常用的SILAC标记都是在精氨酸或赖氨酸上。

化学标记是指通过化学反应对样品直接进行标记。目前,常用的化学标记技术有同位素亲和标签技术(isotope-coded affinity tags, ICAT)^[37],同位素标记蛋白质标签(isotope-coded protein labels, ICPL)^[38], ^{18}O ^[39]等。不同的标记方式对应不同的标记基团与化学反应原理,具体可以参考Bantscheff等人^[40]的综述。

母离子标记的定量算法流程除不需要进行保留时间对齐、交叉搜索和AMT匹配外,与无标定量中基于离子流色谱峰的方法类似。具体来说,在XIC构建时,根据已知的标记质量差,母离子标记定量需要分别构建未标记肽段和标记肽段的XIC,形成一个XIC对。在一个肽段对应的XIC对内部计算未标记肽段和标记肽段相应的定量指标的比值,作为该肽段的丰度比值。蛋白质层面的计算与无标定量的方法相同。

2.2 子离子标记定量

子离子标记^[41]是通过化学反应给肽段添加质量标签,在质谱进行二级碎裂时将质量标签碎裂成不同质量的报告离子,利用不同报告离子的丰度比作为不同样本对应肽段的定量比值。目前最常见的两种子离子标记方法分别是相对和绝对定量的同位素标记技术(isobaric tag for relative and absolute quantitation, iTRAQ)^[42]和串联质量标签技术(tandem mass tag, TMT)^[43]。

iTRAQ试剂由3个化学基团组成,分别是反应基团、报告基团和平衡基团。反应基团特异与肽段的氨基(肽段的N端、赖氨酸的氨基)发生结合反应,有的报道称也可与酪氨酸反应^[44]。报告基团的质量数是不同的,加上相应的平衡基团后,保持iTRAQ试剂总的质量数固定,同时保证不同标记的肽段在色谱中的保留时间相同。由于是等质量的,被不同标记的肽段在一级图谱中表现为一个母离子峰,进一步碎裂时,平衡基团发生中性丢失,报告基团则会产生相应的不同质量数的报告离子,比较报告离子的丰度值,可以得到样本的相对丰度比。目前iTRAQ可以同时标记4个或者8个样本进行标记与定量。TMT与iTRAQ的原理类似,均是通过组合不同的报告基团和平衡基团实现多重定量,此技术目前可以对2个、6个和10个样本进行多重标记。

子离子标记定量与母离子标记定量、无标定量在原理上差别很大(表1)。子离子标记定量计算简便,只需要二级谱图中报告离子的信号强度即可,不需要构建离子流色谱峰。但是,子离子标记也有其自身的局限性,比如易受共洗脱肽段的影响,同时存在丰度比值抑制现象^[45],因此针对具体问题需要采用相应的定量方法。如图4所示,子离子定量流程一般包括以下几步:(i)二级谱图预处理;(ii)报告离子提取;(iii)报告离子校正;(iv)归一化;(v)肽段及蛋白质定量结果计算。其中,在第一步“二级谱图预处理”时,由于报告离子一般都处于质谱检测范围中的低质量区域,相对来说噪声会更高一些,非肽段信号的干扰峰也更多,因此往往需要对这些子离子标记的谱图做相应的去噪和去卷积预处理。预处理之后一般按照设定的质荷比误差容忍度提取已知质量数的报告离子的信号强度。由于不同报告离子带有不同的误差^[46],因此一般都需要对提取出的报告离子按照预先测得的校正系数矩阵做校正,不同的标记方法具有不同的校正系数^[47],这就是报告离子的校正。

归一化是子离子标记定量中比较重要的一步,因为子离子标记的误差不仅来源于样本制备与实验操作,还来源于不同报告离子的误差。一般来说,首先需要对不同报告离子进行归一化,消除同一组实验内部不同报告离子间的系统偏差,然后再对不同实验间的定量结果归一化。Karp等人^[48]采用VSN(variance-stabilizing normalization)算法做归一化,有效降低了iTRAQ标记中的定量误差。通过报告离子

表 1 母离子标记与子离子标记常用方法总结

标记方法	标记原理	主要优点	主要缺点
¹⁵ N/ ¹³ C	蛋白质中掺入 ¹⁵ N/ ¹³ C 原子	标记效率高, 多用于体内标记和植物蛋白质组学	只能标记两个样本, 其标记后蛋白质质量偏移不定
SILAC	蛋白质中掺入同位素标记的氨基酸	标记效率高, 适合细胞标记	成本高, 周期长
¹⁸ O	胰蛋白酶催化下, 多肽 C 端 ¹⁶ O 被置换为 ¹⁸ O	标记操作简单, 成本相对较低	标记效率因多肽性质不同差异较大, 标记的稳定性欠佳, 其质量差不固定(2 或 4 Da)
ICAT	与多肽的-SH 反应从而引入质量标签	标记效率高, 标记多肽易与非标记多肽分离, 操作简单	只能标记两个样本, 只能标记含巯基的多肽, 覆盖率低
iTRAQ	与多肽的-NH ₂ 反应从而引入质量标签	标记效率高, 4/8 重标记	存在定量比值抑制现象; 共洗脱离子干扰; 标记试剂昂贵
TMT	与多肽的-NH ₂ 反应从而引入质量标签	标记效率高, 2/6/10 重标记	存在定量比值抑制现象; 共洗脱离子干扰; 标记试剂昂贵

的校正和归一化, 能够最大限度地减少子离子标记的比值抑制程度. 但是对于共洗脱肽段的干扰问题, 目前主要通过优化质谱实验参数或者生成相应的三级谱图等实验方法进行改进, 仍缺少合适的处理算法. 子离子标记定量的肽段和蛋白质丰度计算方法与前文中母离子标记定量、无标定量的相应方法区别不大.

3 无标定量与有标定量比较

上述各种定量方法不同的原理和特点决定了它们定量效果的参差不齐. 针对不同的研究问题, 若能选择合适的定量方法, 将有利于后续的研究. 本文从定量动态范围、定量准确性、定量重复性、计算复杂度、实验成本、实验操作复杂度、对质谱平台稳定性的依赖程度这 7 个方面评估了无标定量与有标定量中各自方法的特点. 如图 5 所示, 根据已有相关文献的总结, 本文对这 7 个指标依次从 1 到 10 进行打分. 分数越高说明相应指标所代表的程度越高, 例如, 某方法“对质谱平台稳定性的依赖程度”打分越高说明该方法对质谱平台稳定性的依赖程度越高.

在无标定量中, 基于离子流色谱峰的方法要比谱图计数法具有更高的准确性、灵敏性和重复性, 谱图计数法针对中高丰度蛋白质的定量结果与基于离子流色谱峰的方法一致性较高, 但是针对低丰度蛋白质的定量效果较差^[49,50].

在母离子标记定量与子离子标记定量的比较中, 一般认为这两种方法在定量准确性上不分伯仲, 但是子离子标记的定量重复性较优^[51]. 另外, 母离子标记中的代谢标记能够达到很高的标记效率, 但是其

标记周期长, 操作复杂.

无标定量与有标定量相比, 其实验操作相对来说更加简单, 可定量比例更大, 定量结果的动态范围也比有标定量要大, 但是无标定量对于质谱平台的稳定性要求较高, 在较小的动态范围内定量准确性和重复性略逊于有标定量. Arosva 等人^[52]比较了 ¹⁵N 代谢标记与基于 XIC 的无标定量算法, 发现在 10 倍以内的差异中 ¹⁵N 标记更加准确, 而 10 倍以外的丰度变化则是无标定量更准. Liu 等人^[53]则比较了 SILAC 和基于 XIC 的无标定量算法, 结果发现虽然无标定量能够有更高的可定量比例, 但是其蛋白质层面的变异系数高于 SILAC 定量, 重复性不如 SILAC. 该结论与 Megger 等人^[54]在无标定量与 TMT 比较中的结论一致. 本研究组^[55]据此建议无标定量适合于那些以发现为主的研究, 而以验证为主的研究可以考虑有标定量结合质谱选择反应监测技术.

4 绝对定量

目前, 上述的无标和有标定量策略与方法已经广泛应用于以研究同一个蛋白质在不同状态间变化的相对定量研究中, 为差异蛋白质筛选、疾病标记物发现等研究提供了有力的技术支撑. 此外, 以测量蛋白质绝对丰度^[56]为主要目标的绝对定量, 能够准确检测每个细胞中不同蛋白质的绝对量或浓度, 对蛋白质复杂网络研究、相互作用研究以及临床诊断和治疗都具有现实的指导意义. 因此, 蛋白质绝对定量方法的研究对整个定量蛋白质组学以及生命科学的发展意义重大.

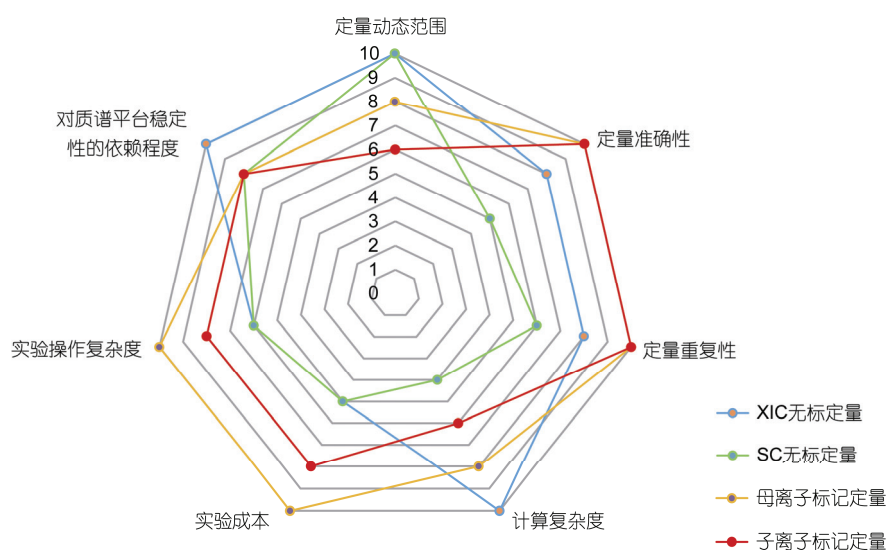


图5 无标定量和有标定量方法特点总结

常用的绝对定量策略往往是在样本中加入已知绝对量的标准蛋白质作为内标, 通过样本和这些内标在质谱中的定量比例关系求得样本中蛋白质的绝对量^[57]. 常用的内标有蛋白质层面的 PSAQ^[58], PrEST-SILAC^[59], FLEXIquant^[60]和 RISQ^[61]等; 也有肽段层面的 QconCAT^[62]和 AQUA^[63]等. 这些内标根据制备原理的不同, 需要在实验操作的不同阶段加入样本中: (i) 上述蛋白质内标需要在样本制备之初加入, 与样本一起混合后酶切; (ii) QconCAT 在样本纯化后, 酶切之前加入; (iii) AQUA 在样本酶切之后加入.

基于内标的绝对定量实验策略能够提供准确的定量结果, 但是由于这些内标在制作上费时费力, 成本较高, 同时也人为增加了样本复杂度, 容易引入实验误差, 因此不基于内标的蛋白质绝对定量方法也逐渐兴起. 在这类方法中, 研究人员希望蛋白质丰度计算值尽可能地接近蛋白质理论丰度, 最后将丰度计算值乘以一个转换系数(如浓度), 即转换为蛋白质的实际丰度. 上文提到的 Top3 法、APEX 法和 emPAI 法均属于此类方法. 另外, Schwanhauser 等人^[64]又提出了 iBAQ(intensity-based absolute quantification)算法, 将一个蛋白质对应的全部肽段定量值之和除以该蛋白质理论酶切肽段数作为表征该蛋白质绝对丰度的指标. Ahrne 等人^[65]比较了这 4 种蛋白质绝对

丰度指标, 认为基于离子流色谱峰的方法(iBAQ, Top3)整体上要比基于谱图计数(APEX, emPAI)的方法更加接近蛋白质理论丰度, 同时对于长度较短的蛋白质, iBAQ 比 Top3 更准确. 最近, Wisniewski 等人^[66]又提出一种新的绝对定量方法: 他们利用真核生物中组蛋白质量与 DNA 质量近似 1:1 的原理, 通过计算组蛋白质的实际质量, 并以组蛋白质的丰度计算值与实际质量的比例为“桥梁”, 计算出样本中其他蛋白质的绝对量.

需要注意的是, 上述绝对定量方法都忽略了一个前提——不同肽段在质谱中的鉴定和定量效率是不同的. 因此直接比较不同的蛋白质和肽段之间的绝对量仍存在不少困难^[67], 绝对定量方法的研究仍有待进一步发展.

5 定量软件及工具

工具是方法的有效载体, 定量方法的蓬勃发展必然也使得相应的定量软件和工具不断涌现. 目前常用的定量软件和工具多种多样, 本文按照不同定量方法类别, 挑选其中的代表性工具进行总结, 分别列举了无标定量、母离子标记定量和子离子标记定量中的常用软件, 如表 2~4 所示, 并简要总结了这些软件的特点、适用的操作系统以及参考文献.

6 定量结果的质量控制

如上所述, 目前针对不同的定量实验策略, 虽然研究人员已经发展了众多的定量算法及工具. 但是, 大部分定量算法和工具并没有考虑定量结果的质量控制, 相应的质量控制方法也缺少系统研究, 这极大地影响了定量蛋白质组学的进一步发展.

要研究定量结果的质量控制方法, 首先要弄清楚定量结果中的误差来源. 对此, Bakalarski 等人^[86]指出定量结果中的误差源于实验操作、定量算法甚至是实验仪器等多方面的干扰, 同时他们认为信噪比(signal-to-noise ratio)可以作为定量结果是否准确的

指标, 也可以近似表征肽段丰度. Piehowski 等人^[87]针对无标定量中的误差来源进行研究, 认为大部分误差来源于样本制备, 而酶切和仪器稳定性的误差并不大. Zhang 等人^[88]分析了有标定量结果中误差产生的原因, 认为误差除了来源于样本制备和酶切外, 还与所选的定量方法有关.

目前已有研究在定量算法研究的同时, 开始关注于定量结果的质量控制. Nefedov 等人^[89]应用中等精度的质谱数据, 使用支持向量机(supporting vector machine, SVM)构建了一个评估 ^{18}O 标记数据定量准确性的模型, 得到了比信噪比更好的效果. MaXIC-Q^[72]提供了关于信噪比、电荷以及同位素峰

表 2 常用无标定量工具列表

软件名称	特点简介	操作系统	参考文献
MsInspect	AMT 匹配, 图形界面/命令行	Linux/Windows	[68]
SuperHirn	AMT 匹配, 与 TPP 兼容	Linux	[9]
SIEVE	Thermo 公司的商业软件, 只支持 Thermo raw	Windows	—
MaxLFQ	MaxQuant 的无标定量模块	Windows	[8]
Census	加权平均求蛋白质丰度比值, 还可以做有标定量	Windows	[69]
LFQuant	非线性可逆的 RT 对齐模型	Windows	[16]
IDEAL-Q	RT 预测与对齐新算法	Windows	[15]
MassChroQ	非线性 RT 对齐模型	Linux/Windows	[70]
TOPP/OpenMS	完整的质谱数据分析流程	Linux/Windows	[11]
DanteR	基于 R 的定量及统计分析工具	Linux/Windows	[71]

表 3 常用母离子标记定量工具列表

软件名称	特点简介	操作系统	参考文献
MaXIC-Q	ICAT 和 SILAC 标记, web-service	Linux/Windows	[72]
XPRESS	ICAT 和 SILAC 标记, 整合在 TPP 中	Linux/Windows	[73]
ASAPRatio	母离子标记, 整合在 TPP 中	Linux/Windows	[74]
MFPaQ	基于 Mascot, ICAT 和 SILAC 标记	Windows	[75]
MaxQuant	质谱数据定性定量的全流程分析, 包括母离子标记定量	Windows	[76]
MSQuant	基于 Mascot, 母离子标记定量软件	Windows	[77]
SILVER	定量质控打分, SILAC, ^{18}O , ^{15}N 标记	Windows	[22]
Census	^{15}N 标记和无标定量软件	Linux/Windows	[69]
UNiquant	仅针对 SILAC 标记, XIC 信号加和作为定量值	Windows	[78]
IsoQuant	仅针对 SILAC 标记	Windows	[79]
pQuant	SILAC 和 ^{15}N 标记	Windows	[80]

表 4 常用子离子标记定量工具列表

软件名称	特点简介	操作系统	参考文献
Libra	整合在 TPP 中	Linux/Windows	[81]
OCAP	小波变换去噪	Linux/Windows	[82]
Isobar	R 包, 开源、免费	Linux/Windows	[83]
WHATraq	包括差异蛋白质筛选	Windows	[46]
Census2	加权归一化算法	Linux/Windows	[84]
MaxQuant	质谱数据定性定量的全流程分析, 包括子离子标记定量	Windows	[76]
IQuant	基于 Mascot, 包括差异蛋白质筛选	Linux/Windows	[85]

簇等特征的过滤条件, 对定量结果进行了质量控制. *msCompare*^[90]在比较各种定量算法与定量模块时考虑了定量算法的质量评估问题, 以内标肽段为参照构建了可靠性打分, 但该方法需要在实验设计中预先加入内标, 增加了实验成本. Chang 等人^[22]在有标定量工具 *SILVER* 中定义了轻重标记的特征峰型的自相关系数等多个指标作为定量结果的质量控制过滤打分, 在不降低定量灵敏性的前提下保证了结果的准确性.

虽然上述相关工作已在定量结果的质量控制方面迈出了第一步, 但是该方向的研究仍处于起步阶段, 深入研究定量结果的质量控制方法依然是个挑战.

7 总结与展望

本文全面总结了目前定量蛋白质组学的相关策略和方法, 对其中的无标定量方法和有标定量方法进行了详细描述和各自特点的比较, 还对绝对定量的相关方法进行了介绍, 最后汇总了相应的定量软

件和工具, 并对定量结果的质量控制方法进行了探讨. 目前, 定量蛋白质组学发展迅速, 实验技术和相应的分析方法也都在不断变化升级, 根据不同的实验目的, 正确选择合适的实验策略和分析方法将有助于后续研究. 因此全面系统地了解这些实验策略和计算方法能够有助于相关研究人员深入理解定量蛋白质组学, 加速该领域的发展. 不过, 目前仍有一些问题尚未得到很好地解决: (i) 绝对定量算法研究仍然不够深入, 如何克服不同肽段在质谱中碎裂效率不同对最终定量结果带来的影响仍然是个难题, 直接从质谱数据中得到的蛋白质丰度计算值与蛋白质实际表达量之间仍存在不小的差别; (ii) 定量结果中的假阳性成因有待归纳和总结, 还需要弄清楚从实验到数据的整个过程中各个环节对定量结果的影响, 一种可行的思路是构建相应的统计学模型估计各部分的定量误差, 从而做到更加精确地定量; (iii) 期望上述分析能够指导定量蛋白质组学的实验设计和分析方法研究, 这将是一项有意义的工作.

参考文献

- 1 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, 422: 198–207
- 2 Kuster B, Schirle M, Mallick P, et al. Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol*, 2005, 6: 577–583
- 3 Srivastava S, Srivastava R G. Proteomics in the forefront of cancer biomarker discovery. *J Proteome Res*, 2005, 4: 1098–1103
- 4 Strittmatter E F, Ferguson P L, Tang K, et al. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom*, 2003, 14: 980–991
- 5 Lundgren D H, Hwang S I, Wu L, et al. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics*, 2010, 7: 39–53
- 6 张伟, 张纪阳, 刘辉, 等. 蛋白质质谱分析的无标记定量算法研究进展. *生物化学与生物物理进展*, 2011, 38: 506–518
- 7 Podwojski K, Fritsch A, Chamrad D C, et al. Retention time alignment algorithms for LC/MS data must consider non-linear shifts. *Bioinformatics*, 2009, 25: 758–764
- 8 Cox J, Hein M Y, Luber C A, et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, 2014, 13: 2513–2526
- 9 Mueller L N, Rinner O, Schmidt A, et al. *SuperHirn*—a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics*, 2007, 7: 3470–3480
- 10 Silva J C, Denny R, Dorschel C A, et al. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem*, 2005, 77: 2187–2200
- 11 Kohlbacher O, Reinert K, Gropf C, et al. Topp—the openms proteomics pipeline. *Bioinformatics*, 2007, 23: e191–e197
- 12 Benjamin A M, Thompson J W, Soderblom E J, et al. A flexible statistical model for alignment of label-free proteomics data—incorporating ion mobility and product ion information. *BMC Bioinformatics*, 2013, 14: 364
- 13 Zhang Z. Retention time alignment of LC/MS data by a divide-and-conquer algorithm. *J Am Soc Mass Spectrom*, 2012, 23: 764–772
- 14 Sandin M, Ali A, Hansson K, et al. An adaptive alignment algorithm for quality-controlled label-free LC-MS. *Mol Cell Proteomics*, 2013, 12: 1407–1420
- 15 Tsou C C, Tsai C F, Tsui Y H, et al. IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol Cell Proteomics*, 2010, 9: 131–144
- 16 Zhang W, Zhang J, Xu C, et al. Lfquant: a label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data.

- Proteomics, 2012, 12: 3475–3484
- 17 Nezami Ranjbar M R, Zhao Y, Tadesse M G, et al. Gaussian process regression model for normalization of LC-MS data using scan-level information. *Proteome Sci*, 2013, 11: S13
 - 18 Callister S J, Barry R C, Adkins J N, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res*, 2006, 5: 277–286
 - 19 Webb-Robertson B J, Matzke M M, Jacobs J M, et al. A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics*, 2011, 11: 4736–4741
 - 20 Sandin M, Teleman J, Malmstrom J, et al. Data processing methods and quality control strategies for label-free LC-MS protein quantification. *Biochim Biophys Acta*, 2014, 1844: 29–41
 - 21 Silva J C, Gorenstein M V, Li G Z, et al. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics*, 2006, 5: 144–156
 - 22 Chang C, Zhang J, Han M, et al. Silver: an efficient tool for stable isotope labeling LC-MS data quantitative analysis with quality control methods. *Bioinformatics*, 2014, 30: 586–587
 - 23 Huang T, Wang J, Yu W, et al. Protein inference: a review. *Brief Bioinform*, 2012, 13: 586–614
 - 24 Zhang Y, Wen Z, Washburn M P, et al. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem*, 2010, 82: 2272–2281
 - 25 Zybilov B, Mosley A L, Sardi M E, et al. Statistical analysis of membrane proteome expression changes in *saccharomyces cerevisiae*. *J Proteome Res*, 2006, 5: 2339–2347
 - 26 Nesvizhskii A I, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics*, 2005, 4: 1419–1440
 - 27 Lukasse P N, America A H. Protein inference using peptide quantification patterns. *J Proteome Res*, 2014, 13: 3191–3199
 - 28 Lu P, Vogel C, Wang R, et al. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*, 2007, 25: 117–124
 - 29 Ishihama Y, Oda Y, Tabata T, et al. Exponentially modified protein abundance index (EMPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*, 2005, 4: 1265–1272
 - 30 Griffin N M, Yu J, Long F, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol*, 2010, 28: 83–89
 - 31 Ong S E, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol*, 2005, 1: 252–262
 - 32 Wu Q, Shan Y, Qu Y, et al. Improved accuracy for label-free absolute quantification of proteome by combining the absolute protein expression profiling algorithm and summed tandem mass spectrometric total ion current. *Analyst*, 2014, 139: 138–146
 - 33 Tabb D L, Vega-Montoto L, Rudnick P A, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res*, 2010, 9: 761–776
 - 34 Ong S E, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 2002, 1: 376–386
 - 35 Wu C C, MacCoss M J, Howell K E, et al. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem*, 2004, 76: 4951–4959
 - 36 Lafaye A, Labarre J, Tabet J C, et al. Liquid chromatography-mass spectrometry and ^{15}N metabolic labeling for quantitative metabolic profiling. *Anal Chem*, 2005, 77: 2026–2033
 - 37 Gygi S P, Rist B, Gerber S A, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, 1999, 17: 994–999
 - 38 Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics*, 2005, 5: 4–15
 - 39 Stewart II, Thomson T, Figeys D. ^{18}O labeling: a tool for proteomics. *Rapid Commun Mass Spectrom*, 2001, 15: 2456–2465
 - 40 Bantscheff M, Lemeer S, Savitski M M, et al. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem*, 2012, 404: 939–965
 - 41 Rauniyar N, Yates J R 3rd. Isobaric labeling-based relative quantification in shotgun proteomics. *J Proteome Res*, 2014, 13: 5293–5309
 - 42 Wiese S, Reidegeld K A, Meyer H E, et al. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, 2007, 7: 340–350
 - 43 Thompson A, Schafer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, 2003, 75: 1895–1904
 - 44 Ross P L, Huang Y N, Marchese J N, et al. Multiplexed protein quantitation in *saccharomyces cerevisiae* using amine-reactive isobaric

- tagging reagents. *Mol Cell Proteomics*, 2004, 3: 1154–1169
- 45 Hultin-Rosenberg L, Forshed J, Branca R M, et al. Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data. *Mol Cell Proteomics*, 2013, 12: 2021–2031
- 46 Zhou C, Walker M J, Williamson A J, et al. A hierarchical statistical modeling approach to analyze proteomic isobaric tag for relative and absolute quantitation data. *Bioinformatics*, 2014, 30: 549–558
- 47 Vaudel M, Sickmann A, Martens L. Peptide and protein quantification: a map of the minefield. *Proteomics*, 2010, 10: 650–670
- 48 Karp N A, Huber W, Sadowski P G, et al. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics*, 2010, 9: 1885–1897
- 49 Tu C, Li J, Sheng Q, et al. Systematic assessment of survey scan and MS2-based abundance strategies for label-free quantitative proteomics using high-resolution MS data. *J Proteome Res*, 2014, 13: 2069–2079
- 50 Krey J F, Wilmarth P A, Shin J B, et al. Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *J Proteome Res*, 2014, 13: 1034–1044
- 51 Li Z, Adams R M, Chourey K, et al. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res*, 2012, 11: 1582–1590
- 52 Arsova B, Zauber H, Schulze W X. Precision, proteome coverage, and dynamic range of arabidopsis proteome profiling using ¹⁵N metabolic labeling and label-free approaches. *Mol Cell Proteomics*, 2012, 11: 619–628
- 53 Liu N Q, Dekker L J, Stingl C, et al. Quantitative proteomic analysis of microdissected breast cancer tissues: comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *J Proteome Res*, 2013, 12: 4627–4641
- 54 Megger D A, Pott L L, Ahrens M, et al. Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochim Biophys Acta*, 2014, 1844: 967–976
- 55 常乘, 吴松峰, 马洁, 等. 基于质谱的选择反应监测技术相关策略和方法的研究进展. *生物化学与生物物理进展*, 2012, 39: 1118–1127
- 56 Bronstrup M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Rev Proteomics*, 2004, 1: 503–512
- 57 Villanueva J, Carrascal M, Abian J. Isotope dilution mass spectrometry for absolute quantification in proteomics: concepts and strategies. *J Proteomics*, 2014, 96: 184–199
- 58 Kaiser S E, Riley B E, Shaler T A, et al. Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat Methods*, 2011, 8: 691–696
- 59 Zeiler M, Straube W L, Lundberg E, et al. A protein epitope signature tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol Cell Proteomics*, 2012, 11: O111.009613
- 60 Singh S, Springer M, Steen J, et al. FLEXIquant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J Proteome Res*, 2009, 8: 2201–2210
- 61 Zinn N, Winter D, Lehmann W D. Recombinant isotope labeled and selenium quantified proteins for absolute protein quantification. *Anal Chem*, 2010, 82: 2334–2340
- 62 Beynon R J, Doherty M K, Pratt J M, et al. Multiplexed absolute quantification in proteomics using artificial qcet proteins of concatenated signature peptides. *Nat Methods*, 2005, 2: 587–589
- 63 Gerber S A, Rush J, Stemman O, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA*, 2003, 100: 6940–6945
- 64 Schwanhaussner B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. *Nature*, 2011, 473: 337–342
- 65 Ahrne E, Molzahn L, Glatter T, et al. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics*, 2013, 13: 2567–2578
- 66 Wisniewski J R, Hein M Y, Cox J, et al. A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics*, 2014, 13: 3497–3506
- 67 Mallick P, Schirle M, Chen S S, et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol*, 2007, 25: 125–131
- 68 Bellew M, Coram M, Fitzgibbon M, et al. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics*, 2006, 22: 1902–1909
- 69 Park S K, Venable J D, Xu T, et al. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods*, 2008, 5: 319–322
- 70 Valot B, Langella O, Nano E, et al. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics*, 2011, 11: 3572–3577
- 71 Taverner T, Karpievitch Y V, Polpitiya A D, et al. DanteR: an extensible R-based tool for quantitative analysis of omics data. *Bioinformatics*, 2012, 28: 2404–2406

- 72 Tsou C C, Tsui Y H, Yian Y H, et al. MaxIC-Q web: a fully automated web service using statistical and computational methods for protein quantitation based on stable isotope labeling and LC-MS. *Nucleic Acids Res*, 2009, 37: W661–W669
- 73 Han D K, Eng J, Zhou H, et al. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol*, 2001, 19: 946–951
- 74 Li X J, Zhang H, Ranish J A, et al. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem*, 2003, 75: 6648–6657
- 75 Bouyssié D, Gonzalez de Peredo A, Mouton E, et al. Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomics study of membrane proteins from primary human endothelial cells. *Mol Cell Proteomics*, 2007, 6: 1621–1637
- 76 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 2008, 26: 1367–1372
- 77 Mortensen P, Gouw J W, Olsen J V, et al. MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J Proteome Res*, 2010, 9: 393–403
- 78 Huang X, Tolmachev A V, Shen Y, et al. UNiquant, a program for quantitative proteomics analysis using stable isotope labeling. *J Proteome Res*, 2011, 10: 1228–1237
- 79 Liao Z, Wan Y, Thomas S N, et al. IsoQuant: a software tool for stable isotope labeling by amino acids in cell culture-based mass spectrometry quantitation. *Anal Chem*, 2012, 84: 4535–4543
- 80 Liu C, Song C Q, Yuan Z F, et al. Pquant improves quantitation by keeping out interfering signals and evaluating the accuracy of calculated ratios. *Anal Chem*, 2014, 86: 5286–5294
- 81 Deutsch E W, Mendoza L, Shteynberg D, et al. A guided tour of the trans-proteomic pipeline. *Proteomics*, 2010, 10: 1150–1159
- 82 Wang P, Yang P, Yang J Y. OCAP: an open comprehensive analysis pipeline for iTRAQ. *Bioinformatics*, 2012, 28: 1404–1405
- 83 Breitwieser F P, Muller A, Dayon L, et al. General statistical modeling of data from protein relative expression isobaric tags. *J Proteome Res*, 2011, 10: 2758–2766
- 84 Park S K, Aslanian A, McClatchy D B, et al. Census 2: isobaric labeling data analysis. *Bioinformatics*, 2014, 30: 2208–2209
- 85 Wen B, Zhou R, Feng Q, et al. IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. *Proteomics*, 2014, 14: 2280–2285
- 86 Bakalarski C E, Elias J E, Villen J, et al. The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J Proteome Res*, 2008, 7: 4756–4765
- 87 Piehowski P D, Petyuk V A, Orton D J, et al. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J Proteome Res*, 2013, 12: 2128–2137
- 88 Zhang G, Fenyo D, Neubert T A. Evaluation of the variation in sample preparation for comparative proteomics using stable isotope labeling by amino acids in cell culture. *J Proteome Res*, 2009, 8: 1285–1292
- 89 Nefedov A V, Gilski M J, Sadygov R G. SVM model for quality assessment of medium resolution mass spectra from ^{18}O -water labeling experiments. *J Proteome Res*, 2011, 10: 2095–2103
- 90 Hoekman B, Breitling R, Suits F, et al. MsCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies. *Mol Cell Proteomics*, 2012, 11: M111.015974

Strategies and Algorithms for Quantitative Proteomics based on Mass Spectrometry

CHANG Cheng^{1,2} & ZHU YunPing^{1,2}

1 State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Engineering Research Center for Protein Drugs, Academy of Military Medical Sciences, Beijing 100850, China;

2 National Center for Protein Sciences (Beijing), Beijing 102206, China

Quantitative proteomics has been one of important areas in omics research. The continuous innovation of experiment strategies and computational algorithms prompts the rapid development of quantitative proteomics. The common quantitative strategies are label-free and stable isotope labeled quantification. There are many quantitative algorithms and tools for each kind of quantification strategy. On the one hand, these computational methods improve the further development of quantitative proteomics. On the other hand, they are being updated continuously with the improvement of the quantitative strategies. Therefore, it is necessary to summarize the advances of strategies and algorithms for quantitative proteomics. In this review, we systematically expatiate upon the characters of the label-free and labeled quantitative algorithms from the view of bioinformatics. Also, this review summarizes the development of absolute quantification methods and lists the common software tools for quantitative proteomics. The quality control methods of the quantification results are discussed as well.

mass spectrometry, quantitative proteomics, quantitative algorithm, label-free quantification, labeled quantification, absolute quantification

doi: 10.1360/N052015-00029