

精确修饰位点谱图库的建立与磷酸化蛋白质组的 DIA 解析

张伟 周岳

赛默飞世尔科技（中国）有限公司

关键词

翻译后修饰；DIA；谱图库；磷酸化；定量蛋白质组学

引言

数据非依赖采集（Data-Independent Acquisition, DIA）是当前最热门的质谱采集技术之一，它以非目标的方式将质量范围分为若干窗口，依次并循环采集窗口内所有母离子的二级碎片^[1,2]。DIA 与 SRM 类似，也是基于子离子（transition）定量，相比传统蛋白质组学定量方法具有更好的选择性和更高的准确度。然而，目前DIA在翻译后修饰分析上仍有较大瓶颈。DIA 依赖于 DDA 建立谱图库，而 DDA 数据在搜库鉴定时，修饰位点定位错误的概率较高，特别是磷酸化修饰发生在常见的 S/T/Y 上，若肽段含有 2 个或以上位置接近的 S/T/Y，位点就容易找错。将含有错误位点信息的鉴定结果作为谱图库，就会导致翻译后修饰 DIA 解析结果的不可靠^[3]。因此，DIA 尚难用于大规模的翻译后修饰样本分析。

针对修饰位点的打分算法使修饰位点的定位更加准确，Proteome Discoverer 软件中整合的 phosphoRS/ptmRS 模块^[4]和 MaxQuant 软件的算法^[5]都可以实现位点可信度（Site Probability）的计算，从而获得可靠的位点定位信息。本文基于上述软件对翻译后修饰 DDA 数据进行位点可信度分析，筛选具有准确位点定位的谱图建立谱图库，导入 Skyline 软件，进而实现可靠的翻译后修饰 DIA 解析。将该流程应用于磷酸化样本 DIA 数据分析，成功提取 6401 条高可信度的磷酸化肽段（ $Q < 0.01$ ），占谱图库肽段总数的 98.4%，其中可用于准确定量的肽段（ $CV < 20\%$ ）占 86.9%，有效解决了翻译后修饰 DIA 定量的难题。

实验条件

实验材料和方法

来源于大鼠组织富集的磷酸化样本，最终上样量为 700 ng/run，分别进行 DDA 和 DIA 采集，每种采集模式重复 3 遍。

色谱条件

纳流高效液相色谱仪：EASY-nLC 1000（Thermo Scientific）

分析柱：纳流 C18 色谱柱（长 15cm, ID 75 μm , 粒径 3 μm ）

流动相：A: 0.1% 甲酸水溶液；B: 0.1% 甲酸乙腈溶液

梯度：0–3 min, 3–7% B; 3–95 min, 7–22% B; 95–113 min, 22–35% B; 113–116 min, 35–90% B; 116–120 min, 90% B

流速：300 nL/min

质谱条件

质谱仪：Orbitrap Fusion（Thermo Scientific）；

离子源：NanoFlex；离子模式：正离子；喷雾电压：1.8 kV；

毛细管温度：275° C；S-Lens RF：60%

DDA：分辨率：一级 120,000@ m/z 200，二级 30,000@ m/z 200；AGC：一级 2e5，二级 5e4；二级 Maximum Injection Time：100 ms；碰撞能量：HCD 30%

DIA：质量范围： m/z 400–1200；窗口：25 m/z （窗口间 1 m/z 重叠，实际 Isolation window 设 26 m/z ）；二级分辨率：30,000@ m/z 200；二级 AGC：1e5；二级 Maximum Injection Time：85 ms；碰撞能量：HCD 30%；每个 DIA 循环之间插入一次一级扫描。

数据处理

基于 Proteome Discoverer 2.0 软件进行 DDA 鉴定、位点筛选和谱图库建立，Proteome Discoverer 1.4 和 MaxQuant 也可以实现相同的工作。

搜库鉴定参数：Uniprot 大鼠蛋白数据库，母离子质量偏差：10 ppm；碎片离子质量偏差：0.02 Da；固定修饰：C 烷基化 (+57.021 Da)；动态修饰：M 氧化 (M+15.995 Da)；S/T/Y 磷酸化 (S/T/Y+79.966 Da)；酶：trypsin；Q 值 (Percolator)：< 0.01；ptmRS 模块：PhosphoRS mode: True; Use diagnostic ions: True

位点筛选和谱图库建立：对 PSM 表格 “Isoform Confidence Probability” 一项进行筛选，保留 ≥ 0.75 的结果，导出 PepXML 格式，并将相应谱图导出为 mzML 格式。PepXML 文件和 mzML 文件同时导入 skyline 建立高可信磷酸化肽段谱图库。

DIA 数据 Skyline 解析：根据 DIA 采集参数设置隔离窗口，将谱图库中所有肽段及相应蛋白作为 Targets，每个肽段选取强度最高的 6 个 b/y 离子进行峰抽提，提峰结果使用 mProphet 进行假阳性评估，控制 Q 值 < 0.01。

实验结果

1. 精确修饰位点谱图库建立与 DIA 解析流程

由于肽段中含有多个可能发生翻译后修饰的位点，如果不对可能发生修饰的位点进行可信度打分，会造成翻译后修饰位点的错误匹配和定位。将错误匹配和定位的翻译后修饰肽段作为谱图库，就会导致 DIA 解析结果的错误。这是目前翻译后修饰 DIA 分析的瓶颈所在。

基于 Proteome Discoverer 的 phosphoRS/ptmRS 算法能够对可能发生修饰的位点进行打分 (Site Probability)，判断定位的准确性。通常认为 Probability ≥ 75 (100 分制) 或 0.75 (1 分制) 的位点定位准确、可靠；具有多个修饰位点的肽段，其所有位点的 Probability 均 ≥ 75 (或 0.75)，则修饰位点及 isoform 唯一确定。通过这一方法，筛选出位点准确可信、isoform 唯一确定的 PSM 即 Isoform Confidence Probability ≥ 75 (或 0.75) 建立谱图库，实现精确的翻译后修饰 DIA 分析。整个流程如图 1 所示。

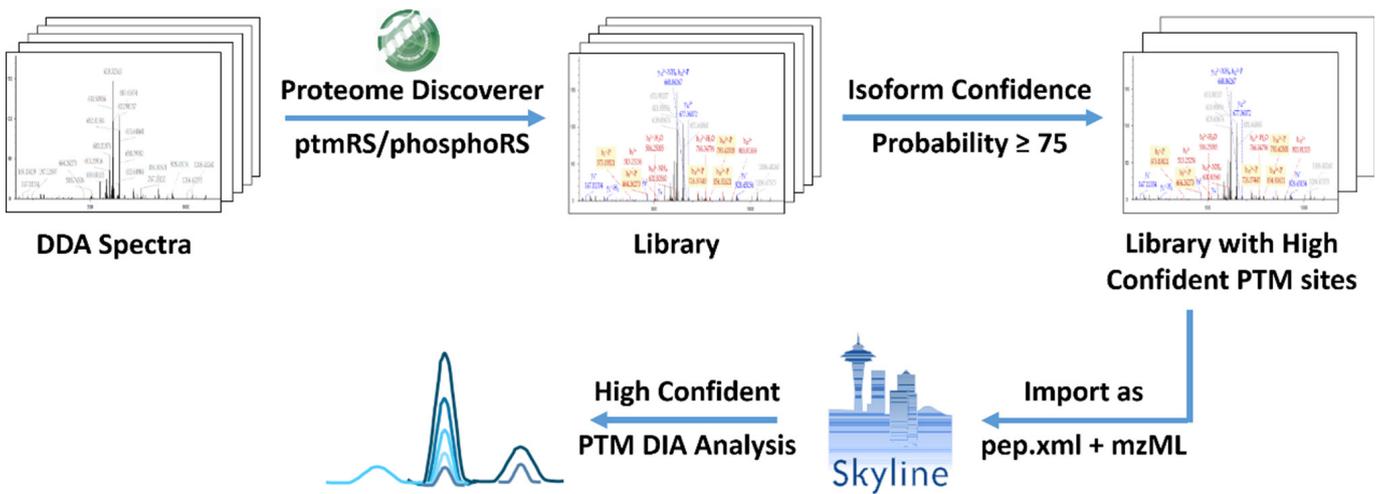


图 1. 精确修饰位点谱图库建立流程图

2. 磷酸化样本的 DDA 鉴定、可信度筛选和谱图库建立

磷酸化样本信息和色谱质谱参数见实验条件部分。3 针 DDA 数据按磷酸化检索流程使用 Proteome Discoverer 2.0 软件搜库鉴定 (S/T/Y+79.966 Da)，并使用 ptmRS 模块对位点打分 (图 2-1)。搜库完成后，打开结果，使用 Filter 功能对 PSM 列表 “Isoform Confidence Probability” 项进行筛选，保留得分大于等于 0.75 的 PSM (图 2-2)。然后，右键选择 “Check All-In This Table”，将符合条件的 PSM 选定。最后，在 “Spectra” 中导出 mzML 格式，在 “To PepXML” 中导出 PepXML 格式，并将两个文件置于同一文件夹中，即可作为谱图库导入 Skyline (图 2-3)。

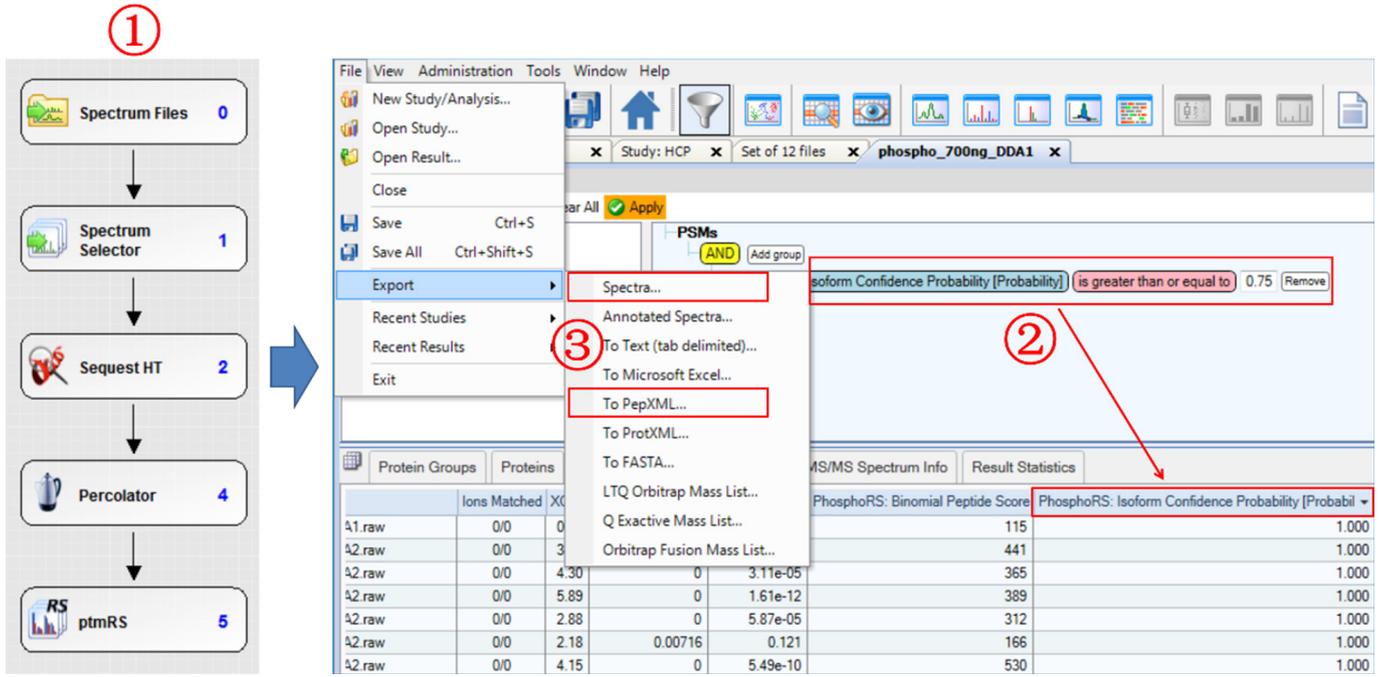


图 2. 翻译后修饰 DDA 数据检索、位点筛选和结果导出步骤

经 phosphoRS/ptmRS 计算得到每个磷酸化位点的打分 (Site Probability)，分数 > 75 (或 0.75) 则位点有确切的碎片支持，定位可靠。搜索引擎没有针对位点打分的功能，位点定位错误概率大。图 3 是一个典型的例子：RFSVTAEGGLTLEQVTDAR 肽段搜库鉴定出包含 1 个磷酸化位点的多个 PSM，3 号位丝氨酸和 5 号位苏氨酸均有匹配到磷酸化发生 (FDR < 1%)。而 ptmRS 的打分结果却分 3 种情况：1) ptmRS 得到唯一可靠的位点，且与搜索引擎得到的位点一致，此时 Isoform 100% 确定 (图 3-A)；2) ptmRS 得到两个位点都有可能，无法确定唯一位点，此时 Isoform 可靠度为 50% (图 3-B)；3) ptmRS 得到唯一可靠的位点，但与搜索引擎得到的位点不一致，此时 Isoform 可靠度为 0 (图 3-C)。最终，只筛选 Isoform 打分 (Isoform Confidence Probability) ≥ 75 (或 0.75)，即所有位点均明确、可靠的肽段建立谱图库。

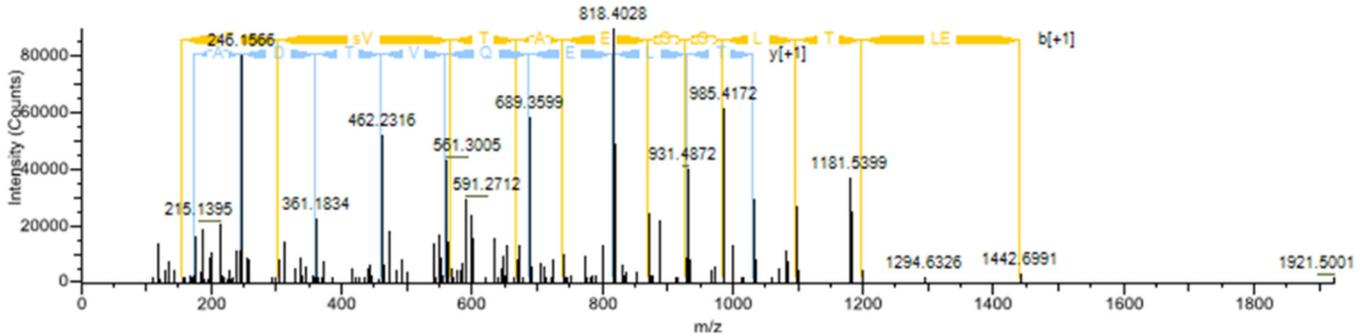
ptmRS结果

100

(A)

搜库结果 R]F]S]V]T]A]E]G]G]L]T]L]E]Q]V]T]D]A]R

Isoform Confident Probability: 100



ptmRS结果

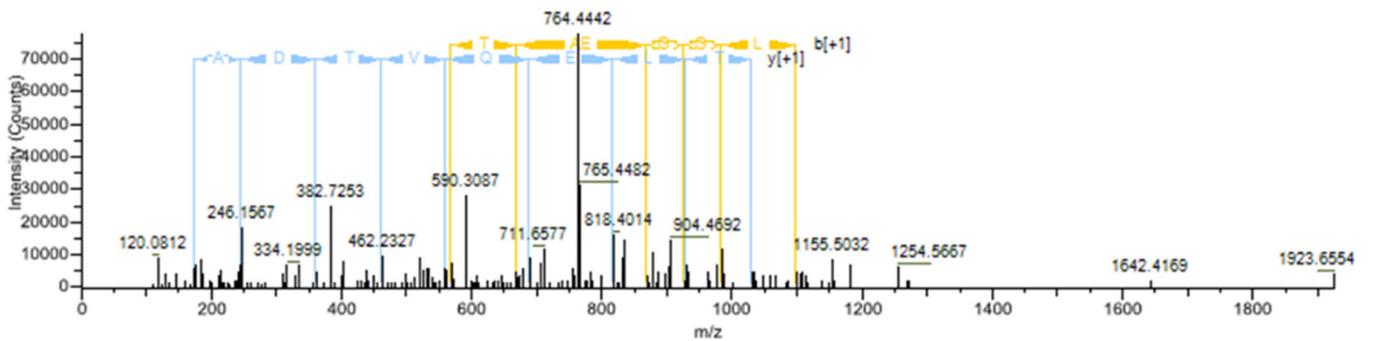
50

50

(B)

搜库结果 R]F]S]V]T]A]E]G]G]L]T]L]E]Q]V]T]D]A]R

Isoform Confident Probability: 50



ptmRS结果

100

(C)

搜库结果 R]F]S]V]t]A]E]G]G]L]T]L]E]Q]V]T]D]A]R

Isoform Confident Probability: 0

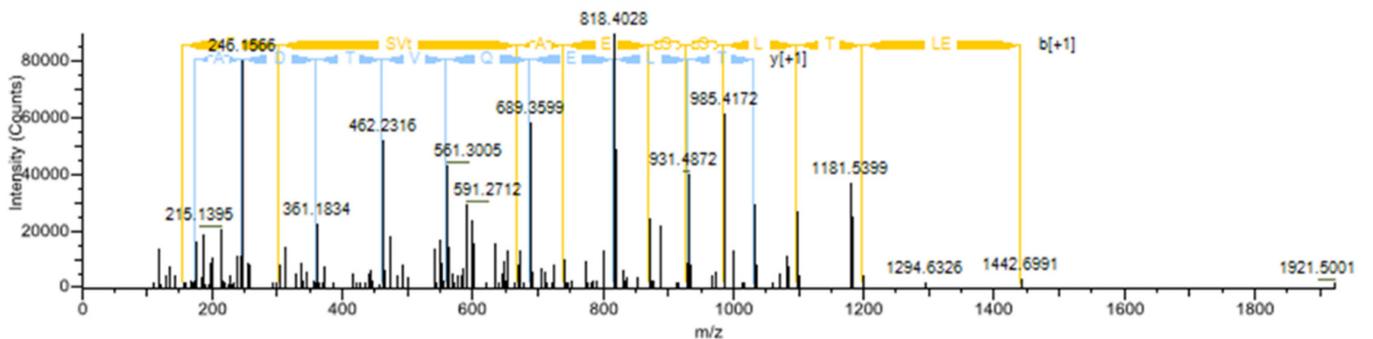


图3. 搜库引擎/ptmRS结果一致、部分一致、不一致三种情况示例

实验共鉴定 56617 张 PSM (FDR < 1%), 经过 Isoform Confident Probability 筛选, 共获得 30558 张精确定位、isoform 唯一的磷酸化肽 PSM (图 4), 结果导出 mzML 和 PepXML 格式作为谱图库。

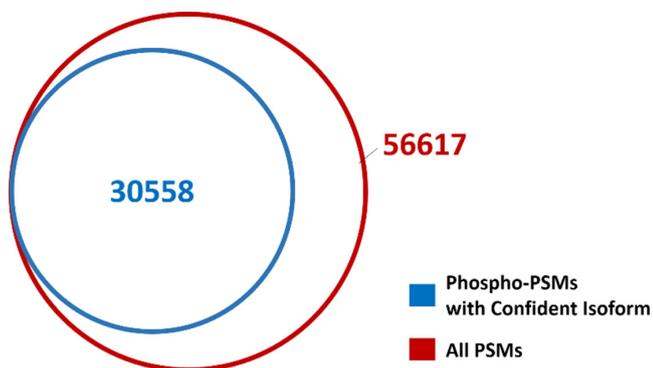


图 4. Isoform Confident Probability 筛选前后 PSM 比较

3. 基于精确谱图库的磷酸化样本 DIA 数据解析

将 mzML 和 PepXML 格式文件导入 Skyline 生成磷酸化肽谱图库, 用于磷酸化肽段 DIA 数据解析。Skyline 将谱图库中所有肽段作为 targets 对 DIA 数据中进行提峰, 每条肽段选取 3 个母离子 (单同位素, 单同位素 +1, 单同位素 +2) 和 6 个响应最高的 b/y 离子。提峰结果使用 mProphet 进行假阳性评估, Q 值 < 0.01 (即 FDR < 1%) 为可信的提峰结果。

结果显示, 从 DIA 数据中提取、定量到 6401 条可信的磷酸化肽, 占谱图库磷酸化肽总数 (6505 条) 的 98.4% (图 5)。磷酸化肽的丰度和离子化效率普遍较低, 本实验如此高的解析成功率表明, 基于 Orbitrap 的 DIA 数据具有极高的谱图质量和出色的灵敏度。

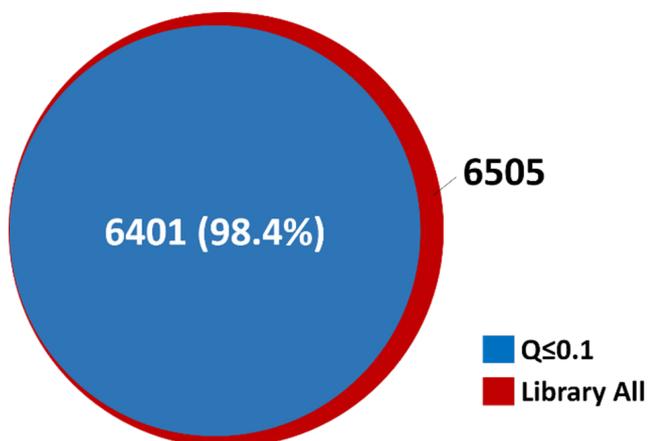


图 5. DIA 可靠解析的磷酸化肽占谱图库总数的 98.4%

进一步对 6401 条磷酸化肽 XIC 色谱峰面积的三针重现性进行统计。将每条肽的母离子和子离子峰面积分别加和, 选取两者中重现性最好 (即 CV 值最低) 的结果用于定量。一般认为, 峰面积 CV 值 < 20% 时定量结果可靠、准确。统计结果显示, 峰面积 CV 值 < 20% 的磷酸化肽占总数的 86.9% (图 6)。如此高的重现性说明 Orbitrap DIA 不仅具有出色的灵敏度, 同时具有优越的重现性, 胜任复杂的翻译后修饰定量。

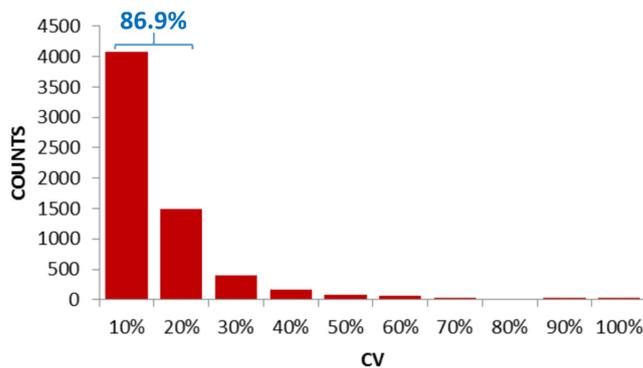


图 6. 磷酸化肽峰面积重现性 (CV 值) 统计

图 7 展示了一个典型的磷酸化肽 DIA 解析结果。该磷酸化肽有 2 种异构体, 分别在 3 号位和 9 号位的丝氨酸上发生了磷酸化。得益于高质量的 DIA 谱图和严格的谱图库建立, 即使这 2 种异构体保留时间非常接近, 也能成功、准确地分辨。DIA 获得的碎片丰度比与谱图库非常接近, 匹配打分的 dotp 值均在 0.95 以上。

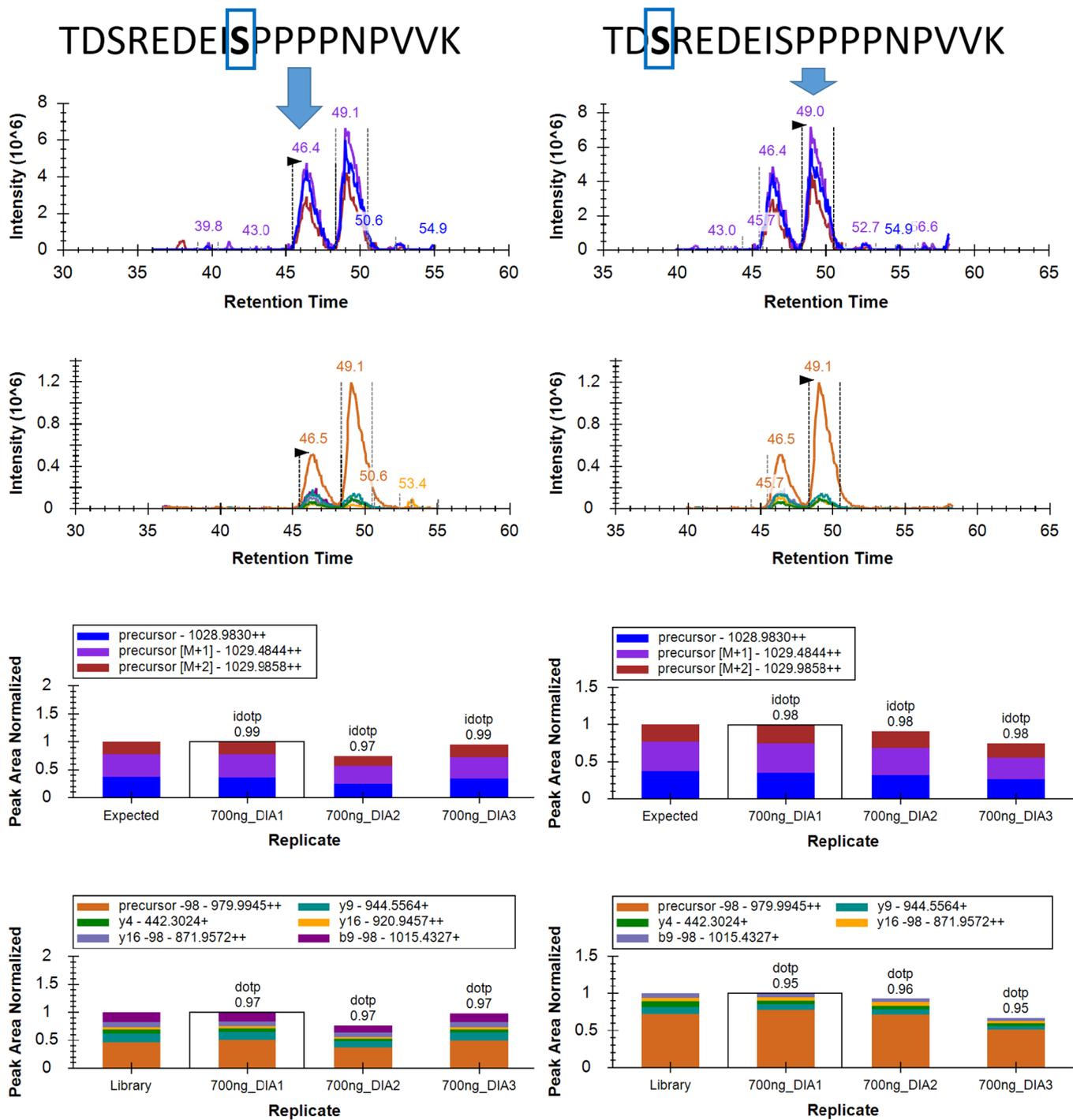


图 7. 磷酸化肽异构体在 DIA 中的分辨和匹配打分 (dotp)

结论

翻译后修饰由于肽段上修饰位点的多重性和不确定性，难以获得可靠的 DIA 结果，这一问题也是 DIA 发展的瓶颈所在。另一方面，phosphoRS/ptmRS 和 MaxQuant 等算法和软件的发展，使翻译后修饰位点的 DDA 鉴定、定位更加准确。本文基于上述算法和软件对磷酸化样本的 DDA 鉴定结果进行位点打分，并筛选出位点定位准确、可靠的谱图建立谱图库用于 DIA 分析，成功从磷酸化样本 DIA 数据中提取 6401 条可靠的磷酸化肽 ($Q < 0.01$)，占谱图库中磷酸化肽总数的 98.4%。其中，86.9% 的肽段峰面积 $CV < 20\%$ ，可用于精确的磷酸化定量。该策略有效解决了翻译后修饰 DIA 定量的难题，证明基于 Orbitrap 超高分辨质谱技术的 DIA 流程兼具出色的灵敏度和优越的重现性，是复杂样本定量特别是翻译后修饰样本定量的最佳选择。

参考文献

- [1] Multiplexed peptide analysis using data-independent acquisition and Skyline, Nat. Protoc., 2015, 10(6): 887-903
- [2] Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues, Mol. Cell. Proteomics, 2015, 14(5): 1400-1410
- [3] MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments, Mol. Cell. Proteomics, 2015, 14(9): 2405-2419
- [4] Universal and confident phosphorylation site localization using phosphoRS, J. Proteome Res., 2011, 10(12): 5354-5362
- [5] MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol., 2008, 26(12): 1367-1372



Orbitrap 组学俱乐部



赛默飞小分子质谱应用技术群

赛默飞世尔科技(中国)有限公司

www.thermofisher.com

全国服务热线: 800 810 5118
400 650 5118 (支持手机用户)

Thermo
SCIENTIFIC

A Thermo Fisher Scientific Brand