

Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology

翻译：肖仕迪 郭韦彤 王越

ppt：肖仕迪 郭韦彤 王越

背景

蛋白质组这个名词在十几年前出现于分析基因组的过程中。是指物种或细胞在特定时期表达的所有蛋白质。

系统生物学需要大量的，在各个分子水平的数据。

质谱法（MS）用电场和磁场将运动的离子按它们的质荷比分离后进行检测的方法。

串联MS（MS/MS）。

质谱仪的软件设定质谱中多肽数目预值，然后分离每一个多肽，在质谱仪中碎片化，测量质谱中的碎片。

蛋白质组学的主要阻碍是技术方面，质谱法的发展改变这种情形。

对于蛋白质的全局测量，基于质谱法（MS）的蛋白质组学就成为了有力的，广泛的方法。在大多数情况下，使用液相色谱法（LC）和高分辨率串联的质谱法大规模地识别和定量多肽类。

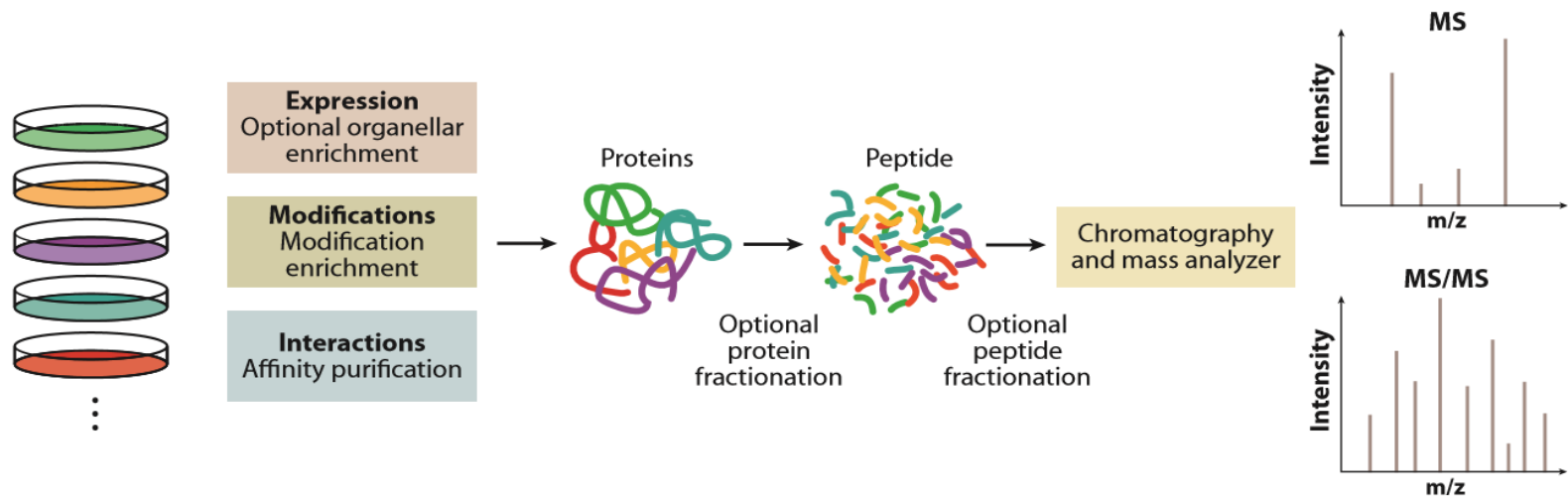
肽和蛋白质识别和量化

在蛋白质组学研究的材料中含有1000到10,000中不同的蛋白质。而基于质谱法的蛋白质组学的主要技术目标是尽量多的准确描述其中蛋白质。

Generic Shotgun Proteomics Workflow (工作流程)

首先，蛋白质降解成多肽（序列特异性的酶：胰蛋白酶）

因为多肽更容易分离和LC-MS分析，MS对小分子量的分子更敏感。



多肽用梯度水或有机溶剂，高效液相色谱法分离，持续一个或几个小时，取决于复杂度。

色谱法和质谱法的分析能产生**MS**数据集和串联**MS**（**MS/MS**）数据，定量的多肽信息。

为了减少复杂度，在蛋白质或多肽的分离中，经典的一维凝胶电泳是必需的，伴随着蛋白质胶内酶切，和在线或离线的阳离子互换。

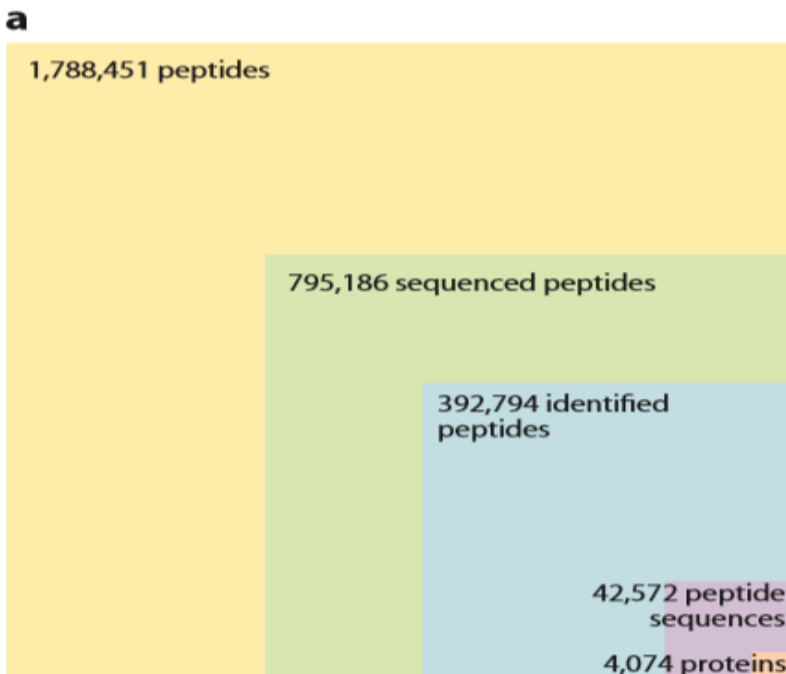
多肽的识别

在多肽洗涤和电喷射的同时，每几秒钟，质谱仪就扫描整个多肽集合，能够使用得到高分辨率的质谱。

多肽的识别率比较低，但现在的高分辨率设备和计算蛋白质组学结合，允许肽序列能明确分配到超过半数的串联质谱。

肽链型蛋白质组学不能直接识别蛋白质，而需要从获得的肽数据重建，这被称为蛋白质推断问题。

一个或两个肽通常足以确定一种蛋白质。然而，区分亚型和确定PTMs,需要尽可能多的肽片段序列覆盖率最大化。

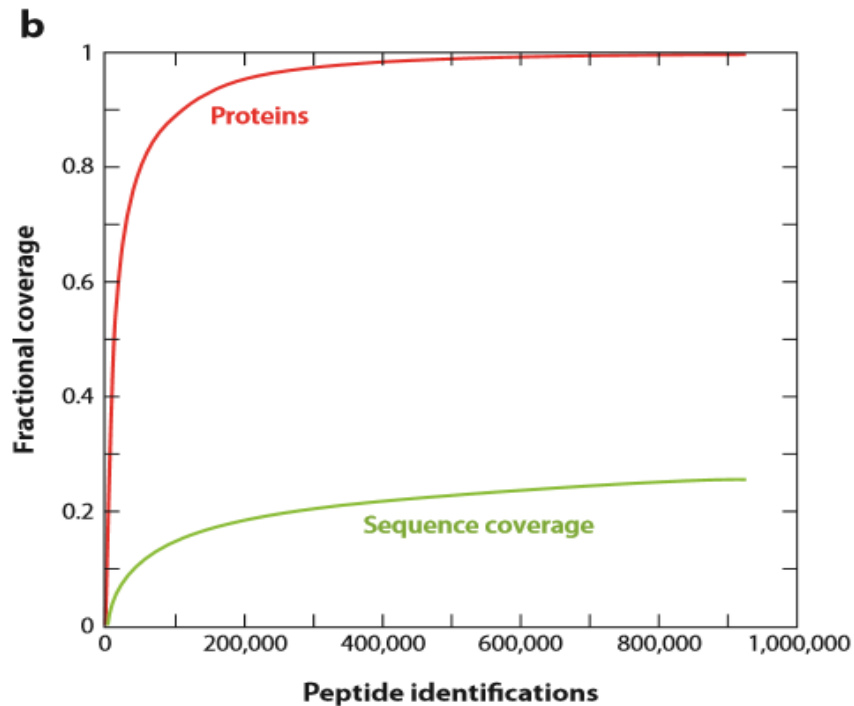


洗脱得到一半的多肽
多肽识别率50%

由于多余测序，不同序列数量更低。

超过**10**个多肽对应一个蛋白质。

不到**10%**的串联的MS已经识别了**80%**的蛋白质。
序列覆盖率逼近**27%**。



蛋白质组学中的定量

定量在MS-based 蛋白质组学中非常重要。

蛋白质组学能确定在混合物中蛋白质的绝对数量和在不同状态下的改变。

而且绝对的定量能提供细胞中蛋白质的拷贝数，这是系统生物学的重要输入。

有两种方法能使蛋白质组定量化，稳定的同位素标记法和无标记法。

不使用同位素的蛋白质定量化准确性低。

而基于同位素的方法是包含特殊的‘重’分子在多肽中，使用化学衍生化或代谢标记。

化学衍生化：

例如，酰胺组的多肽被轻或重（ N^{15} ， C^{13} ， H^2 ）化学元素甲基化。这提供准确的和有利的定量化，而不用重复测量。常见的化学衍生化定量的方法是串联质谱标签和同位素标记相对和绝对定量。

对于代谢标记，细胞在同位素培养基上培养。例如， N^{15} 方法，一个细胞群在只含普通同位素 N^{14} 培养基，另一个细胞群在只含 N^{15} 培养基。利用细胞培养稳定同位素标记技术(SILAC)，精氨酸和赖氨酸含有标记，在细胞复制后每个蛋白质都含有。而且SILAC标记导致容易识别质谱的差异。

代谢标记还有优势，就是样本能够在细胞溶解后直接结合，导致非常高的定量准确性。

靶标分析

一旦感兴趣的蛋白质用上述的技术发现，就要研究它们在不同情况下的行为。将MS测量仪专注于特定的肽，与这些蛋白质有关的，是不错的方法。

常见多肽靶标的方法是（multiple reaction monitoring）MRM，是一种特殊类型的大规模质谱仪（三重四极仪器）用于有选择地记录碎片预定事件和特定的肽。

优点是更高的潜在敏感性和更高的通量，在许多系统生物学应用中有优势。

主要的挑战是假阳性的信号，和处理大量的同位素标记的多肽。

从系统生物学的角度，质谱法蛋白质组学产生三种不同的实验结果或数据。

1.表达蛋白质组学： 决定样本中相对或绝对的蛋白质数目，最大好处是自动将转录后和翻译后水平调控纳入考虑。

2.蛋白质的修饰水平： 能识别和定量成熟蛋白质的一级结构的所有元素，包括翻译后水平修饰（PTMs）。

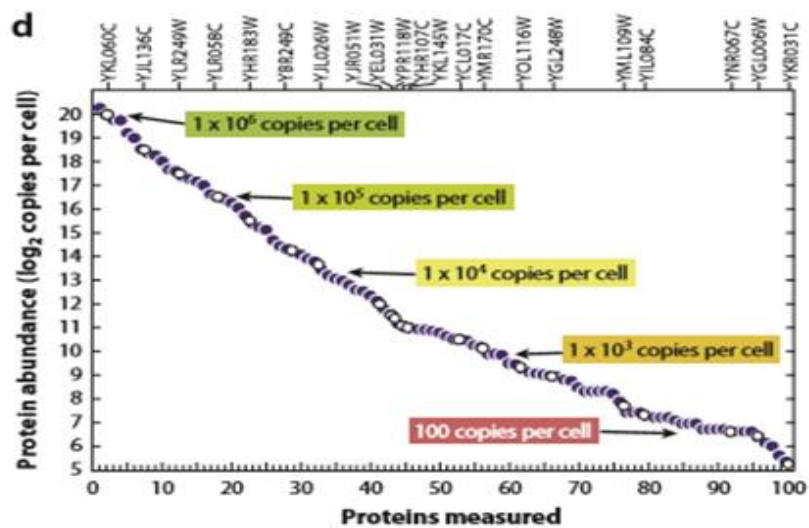
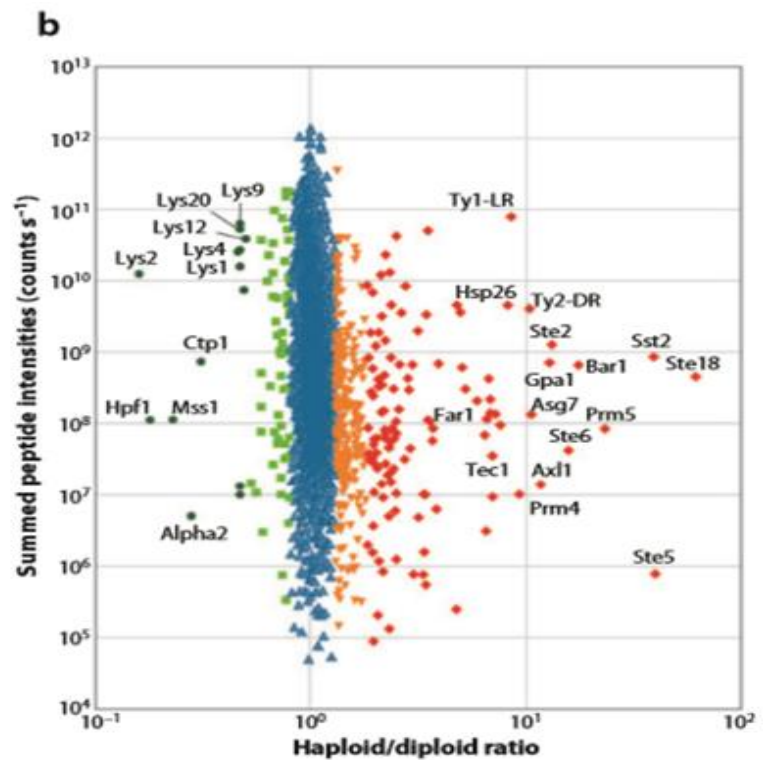
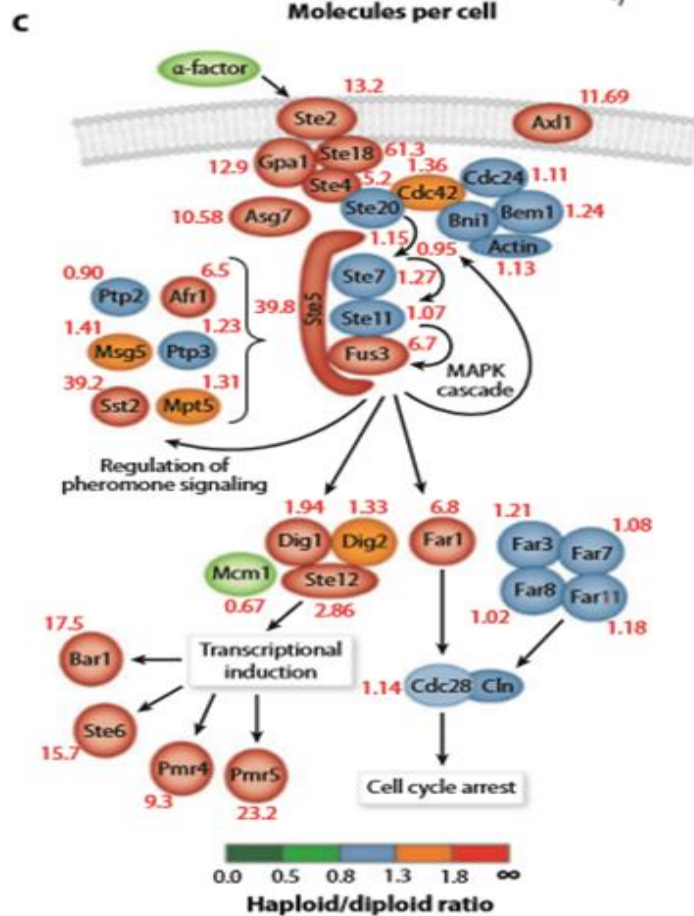
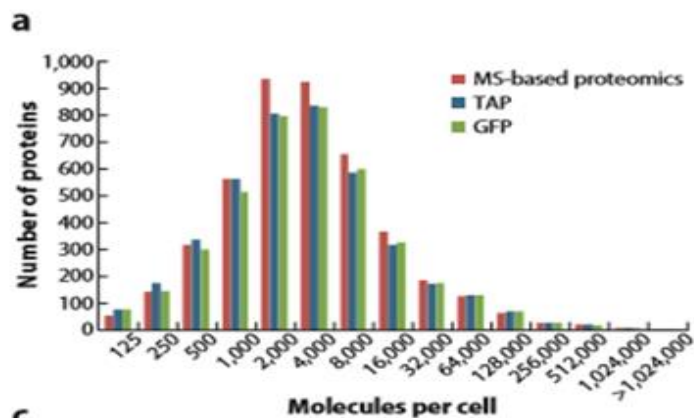
3.蛋白质互作的映射： 对生物网络的贡献不限制于蛋白质和蛋白质的互作，还包含蛋白质与多肽，特殊RNA，DNA序列，和其他小分子的互作。

将这些数据结合起来就能研究生物和医学的问题。

表达蛋白质组学

蛋白质组学测量的是基因表达的终产物，成熟的蛋白，比信号水平相比，与生物学功能联系更紧密。而且，与转录组学不同，表达蛋白质组学能决定，亚细胞间隔和细胞器中的基因产物的水平。

尽管有明显的优势，但是表达蛋白质组学一直被技术困难阻碍。直到鸟枪法蛋白质组学出现，第一个完整的蛋白质组完成定量，在模式生物酵母中，现在已被报道。



现在，在哺乳动物系统中，48小时能够测量大约7000种蛋白质，只需要从样本中获得几千个细胞就足够了。基于在各个蛋白质组学的领域，都有技术的发展，哺乳动物蛋白质组的全面表达分析近在咫尺。

许多蛋白质不仅仅只有一个细胞位置，出现在不同区室，细胞器蛋白质组学是传统的显微镜方法的补充，在细胞生物学中。还有动态的蛋白质组学，不同状态下的蛋白质组。

蛋白质组还能应用于临床的研究。如乳腺癌复发的风险。

翻译后修饰的蛋白质组学

大规模的质谱测量翻译后修饰

PTM分析的主要挑战是修饰蛋白的低丰度，但可以用修饰特异性富集。而且修饰的可能位点和方式的可能性多，不能清楚的识别多肽和修饰。但随着技术的进步，MS-based PTM 分析已经在全局PTM比对方面很成功了。

利用得到的数据能直接测试PTMs可能的效应，研究大规模特定蛋白的特征。例如，进化的保守性，表明磷酸化蛋白质组一般比其他的蛋白质组更保守。

对生物和系统生物学的影响

许多PTMs是共调控的而不是自主的。伴随着敏感开关构造的模式,和许多小型或冗余的方式有助于实现细胞响应扰动。

将PTMs和细胞功能结合在系统的规模下,更容易获得不同扰动的修饰改变的数量信息。

重新思考基因表达和生物控制的观念.

相互作用蛋白质组学

为确定与蛋白关联的功能或代谢通路，研究蛋白互作可能是最有效的手段。如此大量的全蛋白组规模的数据可以揭示细胞组织与功能的很多方面。

定量亲和提纯法复合质谱分析

AP-MS实验的流程也包括5个主要步骤，构建含SBP标签的诱饵蛋白表达载体→目的细胞中过表达带标签的诱饵蛋白→细胞裂解液洗涤→洗脱获得蛋白互作复合物→蛋白互作复合物质谱鉴定

蛋白质组和其他大规模的数据整合

与代谢途径和本体的整合

物理和功能蛋白质的相互作用

蛋白质组和转录组的相互关系

蛋白质组和基因组的结合(SNPs 与细胞功能的联系..)

总结

1. 质谱(MS)的蛋白质组学提供了高分辨率数据成千上万的肽在一个通用的工作流。准确量化通过同位素方法是可能的,但label-free量化方法越来越有吸引力。
2. 第一个完整的蛋白质组已被确认和量化。
3. 蛋白质组中有成千上万的PTMs,但只有一小部分在特定的细胞过程中改变。
4. 蛋白互作能被准确的测量。
5. 蛋白质组能有效的和其他大规模组学数据整合。
6. 蛋白质组学可以提供准确、定量数据为系统生物学的细胞过程。

前景

- 现在蛋白质组学的限制是较低的通量。
- 目前专注于减少测量时间的努力改进的仪器。
- 如果成功的话,这个项目将帮助实现的系统生物学的发展。

thanks