

北京大学

整合科学实践课开题报告

衰老过程中小胶质细胞的单细胞
蛋白质组研究

姓 名： 张广欣

学 号： 2000012272

导 师： 李程

日期： 2024 年 1 月 13 日

中 文 摘 要

蛋白质是细胞生命活动的主要执行者，但由于其无法扩增的特性，获取单细胞蛋白组一直充满挑战。人们常用的色谱-串联质谱（LC-MS/MS）联用技术能够在短时间内获得大量细胞的全蛋白组信息，然而在相当长的时间内其检测限并不足以直接获得单细胞的蛋白质组，也没有成熟的单细胞样本制备方法。目前很多技术路线使用微流控等技术制备可被高灵敏质谱检测单细胞样本，但由于其技术门槛很高而难以普及。

目前笔者所在的李程课题组通过常规仪器配置微升级别的实验体系，结合微量 bulk 细胞样本（carrier、reference）和小分子标签（TMT）获得单细胞蛋白组数据。之后计划使用单细胞蛋白组结合其他组学研究衰老过程中的小胶质细胞，笔者现阶段的工作集中在改进现有的单细胞蛋白组技术，同时利用已有的细胞系蛋白组和转录组数据进行生物信息学分析，技术完善后将研究衰老过程中的小胶质细胞。

小胶质细胞作为大脑的主要免疫细胞之一，负责清除死亡或损伤的神经细胞以及参与调节大脑内的炎症反应。随着年龄的增长，它们的功能和形态将发生变化，从而影响神经系统的健康和稳定。

衰老的小胶质细胞会表现出更强的应激反应能力，产生更多的促炎因子导致持续的炎症状态，从而加剧神经退行性疾病的进展。同时，它们的蛋白质组也会发生改变，影响小胶质细胞清除大脑中异常蛋白质的能力，特别是在阿尔茨海默病中与 β 淀粉样蛋白聚集有关的蛋白质，对衰老小胶质细胞的深入研究有助于理解和应对神经退行性疾病的挑战。

小胶质细胞的异质性在衰老和炎症过程中表现得尤为明显。不同的小胶质细胞亚群可能在衰老和神经退行性疾病的不同阶段中发挥着不同的作用。某些亚群可能在清除脑内积累的蛋白质聚集体中发挥更活跃的作用，而其他亚群则可能更专注于调节炎症反应。因此，在单细胞水平研究衰老的小胶质细胞有着重要意义。

关键词：单细胞蛋白质组, 质谱, 小胶质细胞

Abstract

Proteins are the main executors of cellular life activities, but due to their non-amplifiable nature, obtaining single-cell proteomes has always been filled with challenges. The commonly used chromatography-tandem mass spectrometry (LC-MS/MS) technique allows for the acquisition of whole proteome information from a large number of cells in a short time. However, for a considerable length of time, its detection limit has not been sufficient to directly obtain single-cell proteomes, nor have there been mature single-cell sample preparation methods. Currently, many technical approaches use microfluidics and other technologies to prepare single-cell samples that can be detected by highly sensitive mass spectrometry, but due to their high technical thresholds, they are difficult to popularize.

At present, in the Li Cheng research group where the author is located, a microliter-scale experimental system is set up using conventional instruments, combining minute bulk cell samples (carrier, reference) and small molecule labels (TMT) to obtain single-cell proteome data. The plan is to use single-cell proteomics combined with other omics to study microglia during the aging process. The author's current work is focused on improving existing single-cell proteomics techniques, while also using existing proteomic and transcriptomic data from cell lines for bioinformatics analysis. Once the technology is perfected, it will be applied to study microglia in the aging process.

Microglia, as one of the brain's main immune cells, are responsible for clearing dead or damaged neurons and participating in regulating inflammation within the brain. With aging, their function and morphology will change, thereby affecting the health and stability of the nervous system.

Aging microglia exhibit a stronger stress response capability, producing more pro-inflammatory factors leading to a persistent inflammatory state, thus exacerbating the progression of neurodegenerative diseases. Simultaneously, their proteome also changes, affecting the ability of microglia to clear abnormal proteins in the brain, especially those proteins related to β -amyloid aggregation in Alzheimer's disease. Therefore, in-depth research on aging microglia is helpful in understanding and addressing the challenges of neurodegenerative diseases.

The heterogeneity of microglia becomes particularly evident during aging and inflammation. Different subgroups of microglia may play different roles at various stages of aging and neurodegenerative diseases. Some subgroups may be more active in clearing accumulated protein aggregates in the brain, while others may be more focused on regulating inflammatory responses. Therefore, studying aging microglia at the single-cell level is of significant importance.

KEY WORDS: Single-cell proteomics, Mass spectrometry, Microglia

一、研究背景

目前笔者在实验室中的已有工作主要集中在单细胞蛋白组学的技术建立和优化，由于小胶质细胞分离困难等技术限制，现阶段尚未对小胶质细胞的状态具体展开研究。因此，此处的研究背景将主要梳理与笔者实验室实际工作有关的单细胞蛋白组技术的流变。

1. 蛋白组技术

1.1 常规蛋白组技术以及单细胞蛋白组

蛋白质在细胞中的相对丰度高于对应的转录本，能够更好刻画细胞的生命活动以及与其表型有更好的关联。单个典型的人体细胞中约有 6×10^9 个蛋白质分子，其蛋白质拷贝数目从几千至几百万不等¹，而只有约 3×10^5 个 mRNA 分子²。蛋白质的半衰期显著长于 mRNA^{3,4}，mRNA 到蛋白质会经历转录后调控，翻译后修饰等复杂的生理过程，即使不考虑目前转录组本身的技术缺陷，从这一角度来看，通过分析转录组数据来分析细胞真实的状态的可靠程度也并不高⁵。

分析蛋白组数据将让我们更好地理解细胞的真实生命状态，目前 bulk 蛋白组技术已经相当成熟。有多种技术路线可以实现对细胞中尽可能多的蛋白质进行定量和定性测量，如二维凝胶电泳（2D Gel Electrophoresis）、蛋白质芯片技术（Protein Microarrays）、质谱法（Mass Spectrometry）等⁶。其中质谱法由于其高精度以及对蛋白组覆盖的全面程度而应用最广⁷。

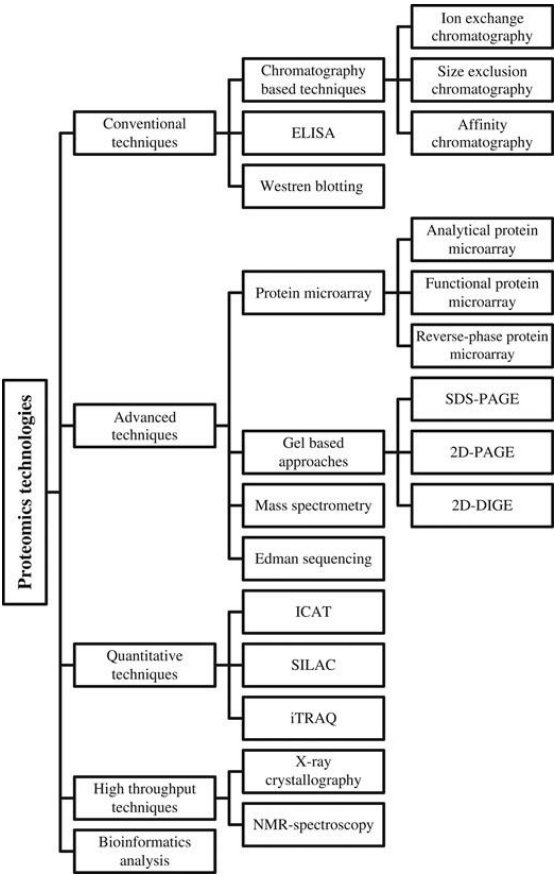


图 1. 各类获得蛋白质组方法⁶

使用质谱法进行蛋白质组信息采集的大致流程如下。首先使用裂解液裂解细胞，从样本中提取蛋白质，之后用碘乙酰胺和二硫苏糖醇打开二硫键，再用胰酶将其消化成肽段，这些肽段可通过质谱采集其质核比信息进行鉴定。目前较常用的策略为电喷雾电离质谱（ESI-MS）。ESI-MS 通过色谱或毛细管电泳等方法对肽段进行分离，随后用高压电离和喷雾化处理，再通过三重四极杆与轨道离子阱对离子进行电离与鉴定⁸（图 2）。

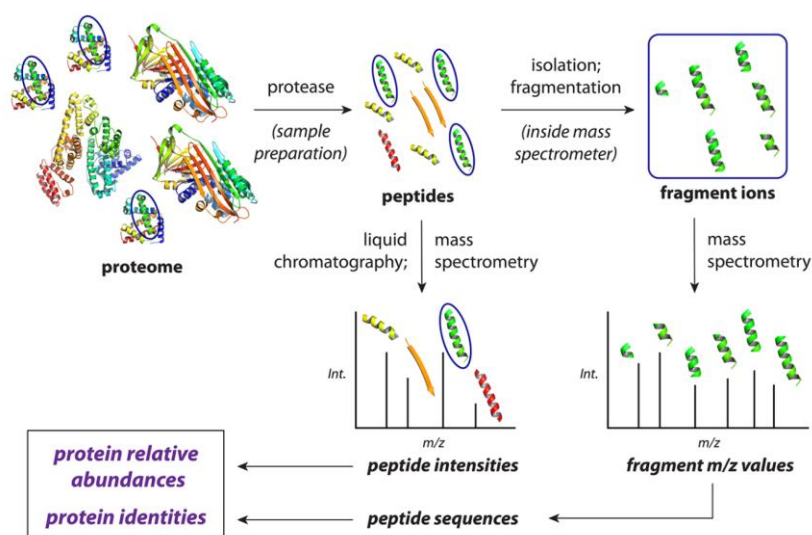


图 2. 简化的典型非靶向自下而上质谱基蛋白质组学实验的概念概述⁹

一般的质谱技术需要较多的细胞样本达到其检测限来保证能检测到尽可能多的蛋白质种类，以实现全蛋白质组范围的分析。但由于质谱采集谱图的原理限制，即难以采集全质荷比谱段的肽段，以及由于长久以来普遍采用的数据依赖型采集手段（DDA, Data Dependent Acquisition）中的鸟枪法会富集到管家基因相关的表达量高的蛋白，而很难采集到微量但对细胞命运影响大的转录因子等。虽然非数据依赖型采集手段（DIA, Data Independent Acquisition）能有效改善这一情况，也在近几年高速发展，但总体并不如 DDA 方案成熟¹⁰。

另一方面，研究单细胞蛋白质组对于揭示细胞内复杂的生物过程和疾病机制至关重要。通过精确刻画单个细胞中的蛋白质表达差异，人们能更好地揭示细胞间的异质性，从而更深刻地理解细胞命运¹¹。

在单细胞水平的蛋白质组学研究中，目前存在的两种主要技术路线为基于抗体和核酸标签的策略和基于质谱的策略：

对于基于抗体和核酸标签的靶向蛋白质组学。2009 年 Scott Tanner 团队创新性地开发了 CyTOF 技术，利用同位素标记的抗体来定量分析单个细胞中的数十种表面蛋白¹²。近年来，有众多后续技术通过结合抗体和可扩增的核酸标签，将蛋白质表达信息转换为核酸序列，以实现单细胞水平上靶向蛋白质组信号的相对定量，包括 CITE-seq¹³、inCITE-seq¹⁴ 和 Histo-Cytometry¹⁵ 等。

然而，抗体-核酸标签的单细胞蛋白质组技术存在一些限制：

首先，由于蛋白质定量高度依赖于抗体，因此能够检测的蛋白质种类受限于抗体的种类（通常为十几种）^{13, 16, 17}，这意味着无法对缺乏特定抗体的蛋白质进行鉴定。

其次，抗体与蛋白的结合可能出现脱靶效应，这将严重影响蛋白质定量的准确性，并产生较高的背景噪音²，大多数技术也仅能对细胞表面蛋白进行定量鉴定，只有少数技术能够对细胞内部和细胞核内的转录因子等蛋白进行定量检测^{14, 18}。

1.2 基于质谱技术的单细胞蛋白质组

高灵敏质谱有效缓解了有限通量、蛋白质有偏鉴定和高脱靶效应等问题。这一技术通过配置纳升至微升规模的实验体系，结合液相色谱联用质谱，根据肽段的相对分子质量可以对其进行分离和梯度洗脱，以此分析和采集质谱数据。单细胞蛋白组的样品前处理一般会用到微流控芯片。

纳升级微流控技术对于单细胞蛋白质组学的样本前处理具有重要意义，是目前单细胞蛋白质组学中的关键技术之一。代表性工作包括 Ryan T. Kelly 与祝莹团队在西北太平洋国家实验室开发的基于玻片操作的 nanoPOTs 和 nanoPOTs (N2)^{19, 20}，以及北京协和医学院的黄超兰、方群和浙江大学合作开发的基于 Oil-Air-Droplet 芯片的单细胞蛋白测定方法¹⁹。这些技术均可以连续处理数十个细胞系样本，在单个细胞中捕获数百至一千种左右的蛋白质。另外，德国 Max Planck 研究所的 Matthias Mann 团队开发的捕获离子迁移率质谱技术 (TIMS) 结合微流控质谱进样系统，可在不同细胞周期阶段的 HeLa 细胞中鉴定出 1000 至 2000 种蛋白质^{21, 22}。

然而，依赖微流控系统和定制化质谱设备的实验流程具有不低的技术门槛，且它们单细胞采集通量和蛋白质种类相对有限。

针对这些问题，美国东北大学的 Nikolai Slavov 教授团队开发了基于常规仪器的高通量单细胞蛋白质组系列技术 SCoPE-MS、SCoPE2 和 pSCoPE，有效克服了行业内其他技术的低通量的技术瓶颈²³⁻²⁵。

其中，本实验室采用的 SCoPE2 (图 3) 利用 384 孔板在微升溶液体系中完成细胞裂解、蛋白质提取和蛋白质酶解等前处理步骤，并通过使用 TMT 标记来自不同单细胞的所有肽段以提高实验通量。同时，通过在单细胞样本间引入微量 bulk 细胞样本 (carrier、reference)，有效提高了低丰度蛋白质的质谱鉴定效率以及实现定量测量。但 SCoPE2 技术有其显著的缺点：

(1) 尽管 SCoPE2 通过 TMT 提高了基于质谱的单细胞蛋白质组技术的通量，但受限于 TMT 的种数，目前仅能一次性对不超过 16 个细胞进行蛋白质定量检测；

(2) 单个细胞中鉴定的蛋白质种类通常在 700 至 1000 种之间^{23, 25}，相比其他技术，它没有在本质上提升鉴定到的蛋白质种数；

(3) TMT 价格不菲，实验流程复杂，也限制了其应用²⁶。

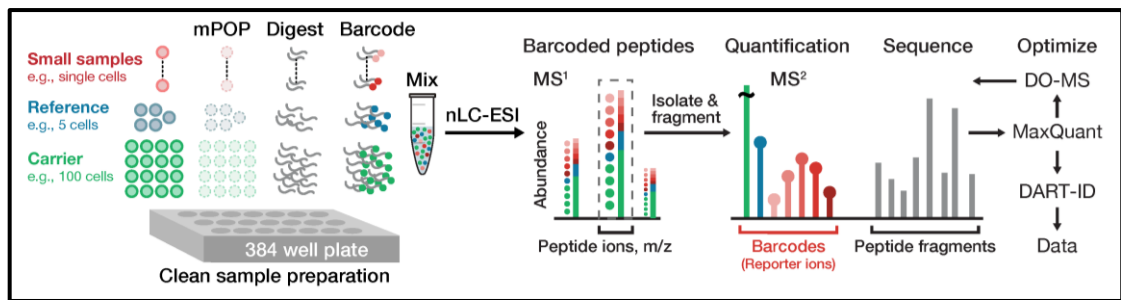


图 3：SCoPE2 的流程图²⁵

此外，单细胞蛋白质组数据普遍存在数据稀疏和组学信息单一的局限性，因此难以仅靠蛋白质组数据精确描绘细胞的完整分子表型。通过数学方法填补稀疏数据（dart-id）并结合现有的单细胞多组学数据进行综合分析，有助于全面理解细胞在重要生物学过程中的基因表达调控机制，对多种组学数据的同时采集与整合也单细胞领域的迫切需求²⁷。

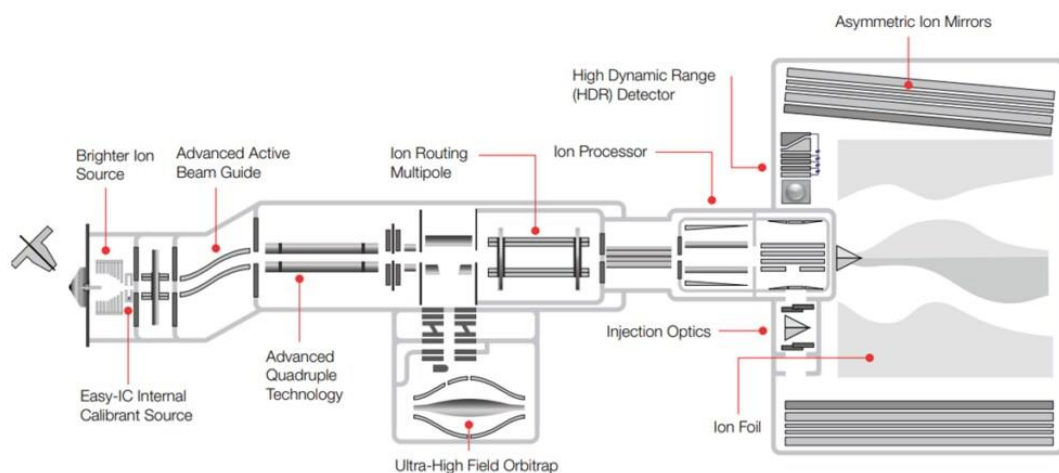


图 4：Astral 结构原理图²⁸

2023 年 6 月，赛默飞公司推出的 Astral 质谱仪（图 4）在单细胞蛋白质组学领域带来了革命性的变革，主要体现在几个方面：

首先，Astral 采用了一种高分辨率、精准质量（HRAM）的仪器平台，结合了 Orbitrap 和创新的 HRAM 非对称轨迹无损分析器。这种独特的设计使得 Astral 在灵敏度和分辨率方面实现了显著提升，从而能够更准确、全面地分析单个细胞的蛋白质组成，实现在真正的单细胞实验中一次鉴定出高达 5000 个蛋白。Astral 的高灵敏度特别适合于检测低丰度蛋白质，如转录因子、激酶和表观遗传调节因子，这些蛋白质在以前的质谱技术中难以检测到²⁹。

其次，Astral 技术在提高单细胞蛋白质组学的通量和蛋白质组深度方面取得了突破。通过使用数据独立获取（DIA）方法，Orbitrap Astral 质谱仪能够在有限的时间内定量更多的肽段，显著提升了蛋白质分析的效率。这一点对于

最大化单细胞蛋白质组学工作流程的生物学影响至关重要²⁸。

2. 衰老过程中的小胶质细胞的变化

随着年龄的增长，小胶质细胞在形态、功能和分子水平上发生显著变化，这些变化对神经系统的健康和疾病状态具有重要影响。在衰老的大脑中，小胶质细胞表现出形态学上的扁平化、树突状突起的减少以及体积变大，这可能影响它们的监视和清除功能^{30,31}。这种形态学变化可能与小胶质细胞的活化状态有关，可能与年龄相关的神经炎症增加有关³²。

在分子层面上，衰老的小胶质细胞表现出炎症标志物的上调，如 IL-1 β 、TNF- α 和 CCL2³³。这种慢性炎症状态可能导致神经元损伤和认知功能下降。此外，衰老小胶质细胞的吞噬功能也受到影响，这可能导致脑内淀粉样蛋白和 Tau 蛋白的积累，进而与阿尔茨海默病等神经退行性疾病的发展有关，但其具体机理尚不清楚³⁴。

最近的研究揭示了衰老小胶质细胞在转录水平上的变化。通过单细胞 RNA 测序技术，研究人员发现衰老小胶质细胞在基因表达上表现出显著的异质性，这可能反映了它们对衰老大脑中多种不同刺激的反应³⁵。

这些发现为理解衰老过程中小胶质细胞的功能变化提供了新的视角，并为开发针对衰老相关神经疾病的治疗策略提供了可能的靶点。

二、拟解决的关键科学问题

1. 如何有效分析和解读通过实验获得的单细胞蛋白组数据，以及如何将单细胞蛋白组的单模态数据与其他组学数据进行联合分析？

单细胞蛋白组数据具有高度的稀疏性，这给上游的谱图的有效匹配和下游的分析造成了巨大挑战，且单一组学只能反映生物系统某一侧面的信息，而多组学联合分析将有助于人们更好理解生命系统。

2. 如何建立高效可行的单细胞蛋白组技术？目前的单细胞蛋白组技术制备流程复杂，实验成本极高，且从单细胞中捕获重要的低丰度蛋白的效率不高。

3. 如何在单细胞蛋白组层面解析衰老过程中的小胶质细胞的异质性，鉴定出与疾病相关的信号通路？目前在转录组层面可以对小胶质细胞进行细致的分群，若能在单细胞蛋白组层面鉴定出这些细胞亚群，将这些差异表达的表型与小胶质细胞的生理状态和功能建立联系，将有助于我们理解疾病底层的分子机制。

三、研究内容

1. 对于已有的单细胞蛋白组数据进行全流程分析，熟悉生信处理流程。单细胞蛋白组数据分析没有如 Seurat 那样极其成熟简化的分析流程。熟悉 R 语言和对已有的单细胞数据进行分析，并结合公共数据进行不同组学数据的整合分

析。

2. 在已有范式下改进实验流程，提升单细胞蛋白组技术的性能，或者切换研究范式，充分利用各种方式建立起简单可靠高效的单细胞蛋白组技术的实验流程：关于改进原有的实验范式，已知原有的冷热交替法裂解获得 bulk 体系的蛋白效率不高，尝试使用不同的裂解液，检测不同裂解液能否有效提升蛋白的释放效率。同时也可以考虑跳出已有范式的桎梏，设计更简单直接的样本制备方法，为即将到来的新一代质谱仪（Astral）做好准备。

3. 从获取大脑不同脑区的 bulk 蛋白组过渡到获取小胶质细胞的 bulk 蛋白组，再发展到获取小胶质细胞的单细胞蛋白组，并结合单细胞转录组或染色质开放性组等进行联合分析。结合科学问题，购买老年小鼠或者从合作者实验室借取模型鼠，利用显微解剖的方法进行脑区分离，再提取小胶质细胞进行后续实验分析。

四、研究目标与方案设计

1. 单细胞蛋白组的生信分析

（1）方案设计

对实验室前期工作产生的单细胞蛋白组数据进行生信分析。上游处理使用 Maxquant 解析谱图，使用基于贝叶斯方法的 dart-id 提升数据质量，再使用 SCP 和 qfeature 等 R 分析工具进行质控和生信分析。结合转录组数据和一些公开的 CITE-seq 数据，利用 Seurat 等单细胞分析工具进行转录后调控分析。

（2）预期结果

充分分析已有的数据，从目前实验室已有的单细胞蛋白组数据中尽可能多地挖掘其蛋白组层面的信息。在这一过程中熟悉 linux 操作、其中的各种工程化操作、基于 R 语言的分析流程从公共数据库下载相关的生物信息学数据以及单细胞转录组的流程，并分辨不同基因受转录后调控的影响。

（3）可能碰到的困难的以及可能的解决方法

由于由 SCoPE2 产生的单细胞蛋白组数据高度复杂，其各类参数设置以及还有 TMT 标记，使得其在 linux 版的 Maxquant 上的搜索过程变得特殊而复杂，后续产生的数据也与一般蛋白组数据的处理流程不同。使用基于 linux 系统的服务器处理可以做到高效并行，但这一基于命令行的处理范式又是复杂且充满工程上的障碍，配置 R 环境充满各种冲突和报错，容易消磨巨量时间。

解决方法是多元的，可以通过寻求网络论坛资源，chatGPT 和官方教程寻找报错的解决方案，也可以向组内有着丰富生信分析经验的师兄师姐寻求直接经验。

2. 对单细胞蛋白组方案的改进

（1）方案设计

想要对于单细胞蛋白组的实验方案进行改进，其出发点是对这一套流程体系足够熟悉，因此需要按照已有的实验指南做一遍单细胞蛋白组实验。第一次

实验可以先用之前已有的部分中间产物，之后从养细胞开始从头进行全流程联系，再根据已有的经验判断需要从哪个部分改进现有的试验方案。

（2）预期结果

熟悉细胞培养等基本的湿实验操作，对于实验室的湿实验规范有清晰的了解。对基于 SCoPE2 的单细胞蛋白组流程充分理解，之后能根据对于实验范式的了解对现有方案进行改进。

（3）可能碰到的困难的以及可能的解决方法

单细胞蛋白组实验在课题组内并不是常规实验，上一批实验中间样本产生时间较长，其质量并不能保证；实验流程长且复杂，每次实验都要经过预先协商才能和他人才可能达成合作，需要很多的预先交流；实验室内做湿实验的成员少，遇到问题难以获得即使有效的帮助；自己掌握实验的全链条固然是必要的，但长远看适当的分工是有意义的。

这些问题的解决首先是将实验的每个流程都做一遍，一旦出现问题，回溯和纠错也较为容易，并可以排除其他非个人实验因素的干扰；每次实验前都要经过细致的预先规划，和各方在前期充分交流，以保证实验的顺利进行；对于实验中的每个阶段都有当时需要聚焦的任务，对于不重要的部分可以进行适当分工，比如细胞培养。

3. 从脑蛋白组到小胶质细胞的单细胞蛋白组

（1）方案设计

订购小鼠，取脑分离脑区后使用 bulk 蛋白组的实验流程获得不同脑区的蛋白组，初步建立起 bulk 蛋白组的实验流程体系。

建立稳定的 bulk 蛋白组的流程体系后，将体系逐渐微量化 (bulk 质谱每次上机只需要 5 微克蛋白，约 2 万个细胞)，再过渡到极微量体系的单细胞样品制备。

之后可以使用 Astral 进行 DIA 的数据采集，使用无标的方法直接得到微量细胞的质谱信息。该方法建立完备后，可以利用其研究脑细胞的异质性。

（2）预期结果

获得各个脑区的完整蛋白组数据，并掌握对小胶质细胞的分离和培养方法。在各个尺度得到稳定可重复的 bulk 蛋白数据，并建立对 bulk 蛋白以及使用 DIA 采集的 bulk 蛋白的分析流程，之后将这流程使用于微量脑组织以至于单个脑细胞，来研究老年小鼠脑中小胶质细胞的异质性。

（3）可能碰到的困难的以及可能的解决方法

与上一部分类似，器材的缺乏以及指导的缺失都是直接显然的问题。

可能的解决方法为向其他有丰富实验经验的实验组寻求合作，哪怕是间接的指导有时也是重要的，实验的基本操作和思想基本是相通的。组内实验经验丰富的师兄师姐的加入合作也尤为重要。

五、初步研究结果

1. 单细胞蛋白组的生信分析

1.1 SCoPE2 数据质控分析

在 Xcalibur 软件中，查看正常流程生成的谱图，检查一系列关键特征的出现，初筛样本质量。

色谱图在 20 分钟左右峰值出现，而信号采集在大约 70 分钟左右基本结束。对于 MS2 的谱图，观察到信号最高的是 126 通道 (carrier)，次高的是 127 通道 (reference)，而其余通道的信号强度约为 127 通道的 1/5 到 1/3 (考虑到 carrier 的裂解效率较低)。这种差异是因为 TMT 126 标记了 carrier，而 127N 标记了 reference，而 127C 和 128N 由于信号串扰问题未被使用。而 128C 到 134N 通道则标记了单细胞，它们的混合比例大致为 100: 5: 1 (图 5A)。

对 PSM 矩阵进行质控，剔除可能来自污染或未在数据库中匹配到的肽段。计算所有单细胞通道中 PSM 相对于 carrier 细胞通道中的平均丰度比值，也就是单细胞报告离子强度与 carrier 细胞报告离子强度的平均比值，简称为 MeanSCR。排除那些 MeanSCR 值超过 10% 的 PSM (图 5B)。在每个批次中，使用 reference 样本作为内参进行批次间单细胞样本的相对定量分析 (图 5C)。同时考察 MedianRI，以排除那些肽段信号强度明显不符合 Carrier/Reference 通道细胞比例的单细胞样本 (图 5D)。

处理所有 PSM 对应的肽段时，整合并去除重复的肽段。为刻画每个肽段的相对丰度，选择每个肽段在 PSM 中的中位数强度，并构建每个批次对应的肽段丰度矩阵。在肽段特征层次上，为评估单细胞样本中蛋白质的肽段丰度变异性，计算每种蛋白质的肽段相对丰度的中位变异系数 (MedianCV)，排除那些 MedianCV 值超过预定阈值 (0.4) 的单细胞样本，因为较高的 MedianCV 指示可能存在蛋白质定量的不稳定性 (图 5E)。

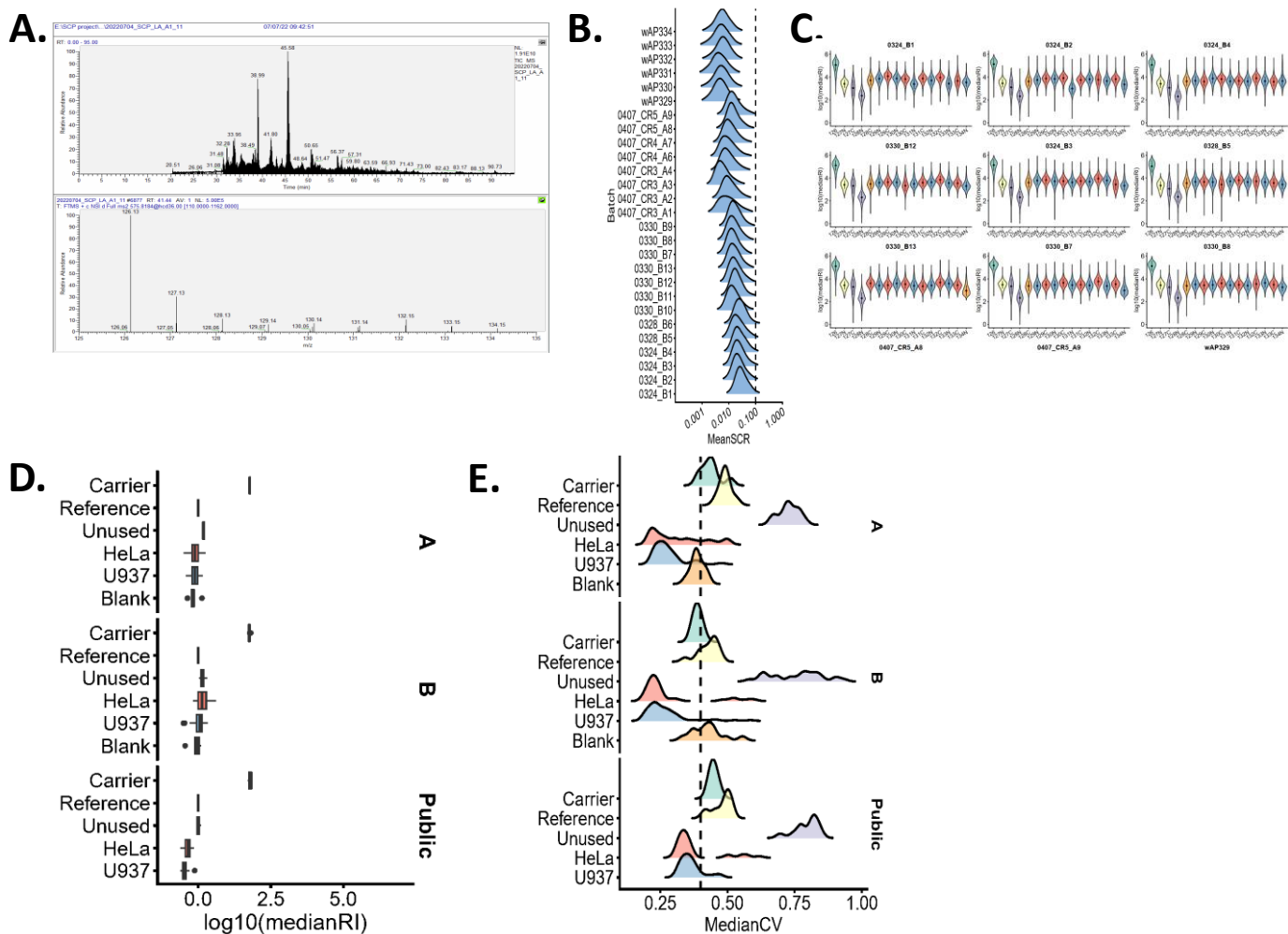


图 5：蛋白组数据质控分析

- (A) 由 SCoPE2 方案产生的色谱以及质谱数据
- (B) 不同批次样本 MedianCV 分布
- (C) 各通道信号的原始信号统计，由 reference 定量（共 33 个，此处展示 9 个）
- (D) 单细胞样本 medianRI 分布
- (E) 不同批次的 medianCV 分布

1.2 单细胞蛋白质组数据集的构建

经过上述质控流程，对结果进行统计分析，合并两轮实验（A、B 组）和 SCoPE2 公共数据，结果显示实验室产生的数据与已发表的公共数据相当（图 6A）。

进一步比较本研究数据与公共数据中的蛋白质表达，发现 HeLa 细胞和 U937 细胞在单个细胞蛋白质表达水平上存在显著差异，表明这两种细胞类型具有不同的蛋白质表达特征（图 6B）。

在公共数据中，每个 HeLa 和 U937 细胞平均能够检测到 500 多种蛋白质，而本研究中单个细胞的蛋白质鉴定数量明显多于公共数据，约为 800~1000 个（图 6C）。

通过对两次实验产生的蛋白质种类与公共数据进行联合分析，共鉴定出了 1839 种不同的蛋白质，超过公共数据库中的数据（A 批次 $n=1630$ ，B 批次 $n=1340$ ，公共数据 $n=998$ ）。与公共数据鉴定的蛋白种类相比，超过 83% 的蛋白种类是相同的，不同批次之间的一致性也很高（图 6D、E、F）。

进一步对高表达的蛋白种类进行基因本体论 (GO) 分析 (图 6H), 发现 HeLa 细胞中高表达的蛋白与蛋白质折叠、细胞质翻译、细胞膜修复等相关。而 U937 细胞高表达的蛋白与蛋白质复合物组装的调节、双链断裂修复通过断裂诱导复制、胞质分裂相关, 这些结果与预期一致。

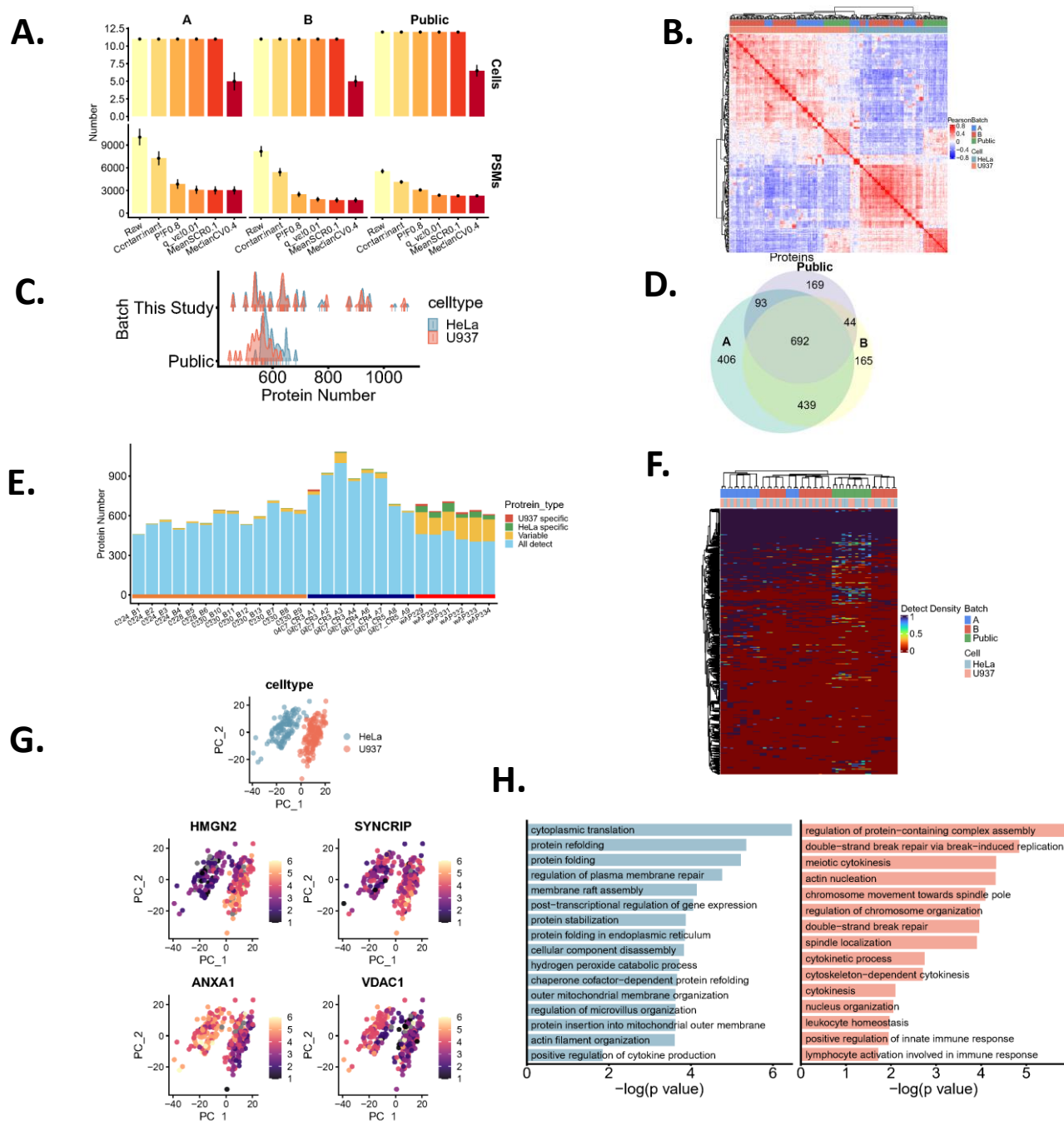


图 6：单细胞蛋白质组数据的构建与比较分析

- (A) 每组质量控制筛选与最终有效保留 PSM 数目
- (B) 两个实验组 (A、B) 和公共数据的单细胞蛋白组热图
- (C) 本研究鉴定到的蛋白中类与公共数据比较的韦恩图
- (D) 各通道信号的原始信号统计，由 reference 定量 (共 33 个，此处展示 9 个)
- (E) 各批次的蛋白质种类以及数目分布
- (F) 样本中不同细胞类型中同一蛋白鉴定频率
- (G) 两种单细胞蛋白组数据的 PCA 以及 marker 蛋白的 feature plot
- (F) HeLa 和 U937 高表达蛋白的基因本体论 (GO) 分析

1.3 将 SCoPE2 数据和转录组数据联合分析

在相同培养条件下，对 HeLa 和 U937 细胞进行单细胞测序，将获得的单细胞转录组数据与单细胞蛋白组数据整合，形成一个矩阵，其中行为基因名称，列为 1018 个单细胞的单细胞转录组和单细胞蛋白组的表达量。对这两种数据进行统一的归一化和标准化，以确保它们具有可比性，发现降维后它们的分布模式不一致 (图 7A、B)。

计算每个基因的转录组表达分布与其他所有基因表达分布的皮尔逊相关系数 (r_i)，以及相应蛋白质的表达分布与所有基因表达分布的皮尔逊相关系数 (p_i)。通过分析 r_i 和 p_i 之间的相关性 (图 7C)，能够探索基因表达在转录层面和翻译层面的协调性。这个分析可以帮助理解基因表达在不同层次之间的关联和调控机制。

通过对相关系数 r_i 和 p_i 的分析，可以将基因分成两个簇，一个受到转录后调控效应相对不强，另一个受到转录后调控效应相对较强。对这两个簇内的基因进行进一步分析绘制热图，发现 cluster I 中的基因所对应的蛋白组和转录组的表达模式类似，而 cluster II 则相反 (图 7D)。

从这两个基因簇中提取了相关系数绝对值前 10 的基因进行 GO 分析 (见图 7E、F)。结果显示，Cluster I 中的基因主要富集在生物正向过程、免疫过程和对刺激的响应过程，表明这些基因可能在维持细胞正常功能和应对外界刺激中发挥重要作用。而 Cluster II 中的基因富集于生物负向过程、定位信号以及稳态维持等过程，暗示这些基因可能在细胞的负反馈调控和稳定性维持中扮演关键角色。

对 Cluster I 中的基因进行分子复合物检测 (Molecular Complex Detection) (图 7G)，结果显示它们所在的复合物在细胞的基本生物化学过程中扮演着核心角色，特别是在转录和核糖体翻译相关的过程中。这意味着这些基因直接参与了 DNA 到 RNA 的转录过程，以及随后的 RNA 到蛋白质的翻译过程，这是细胞功能和表型表达的基础。

对于 Cluster II 中的基因，通过对 GO 分析结果的词条进行关联分析，得到了彼此连接的网络图 (图 7H)。这些发现表明 Cluster II 中的基因在细胞信号传导、代谢调节和细胞生理状态的平衡中扮演着重要的协调作用。通过关

联分析构建的网络图揭示了这些基因之间的相互作用和协同作用，突显了它们在维护细胞稳态和响应环境变化中的集体效应。这些基因不仅在单个水平上发挥作用，而且通过相互作用形成复杂的调控网络，共同参与细胞的多种重要功能。这些发现有助于深入理解基因之间的相互作用和调控机制，进一步揭示了细胞功能的复杂性和多样性。

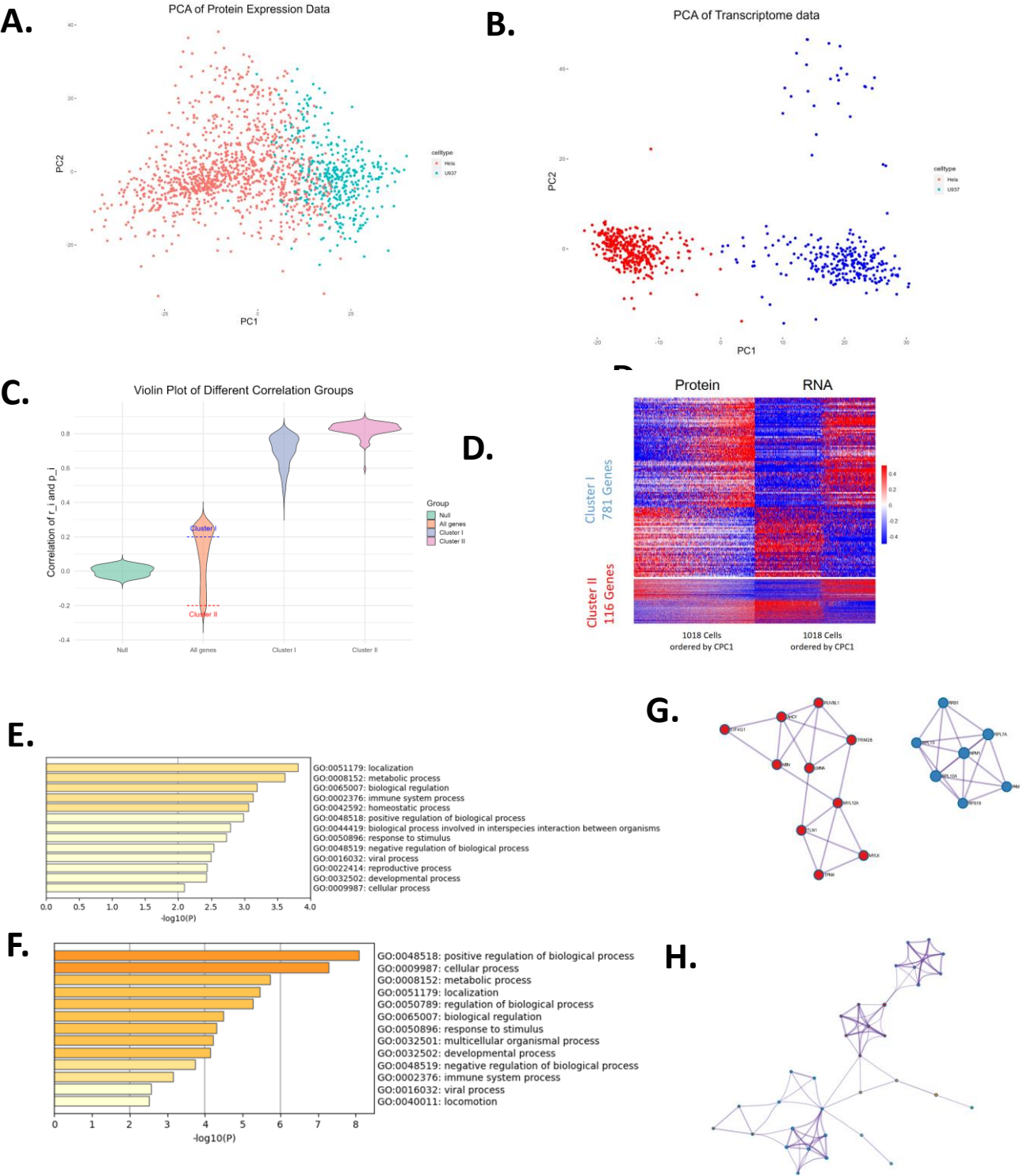


图 7：单细胞蛋白组 (SCoPE2) 与单细胞测序得到转录组数据进行联合比较分析

(A) 由蛋白表达量进行的 PCA 分群

(B) 由转录组表达量进行的 PCA 分群

(C) r_i 和 p_i 之间相关性分布的研究，其中 r_i 是第 i 个 RNA 与所有其他 RNA 之间的成对相关性向量， p_i 是第 i 个蛋白质与所有其他蛋白质之间的成对相关性向量。零假设分布对应于随机排列 RNA 和蛋白质的顺序。所有基因的相关性分布的两种模式被用来定义基因簇 1 和基因簇 2，之后仅在来自基因簇 1 或基因簇 2 的基因空间内重新计算 p_i 和 r_i 之间的相关性，并作为独立的分布显示

(D) 基因簇 1 中的基因在 RNA 和蛋白质水平上显示出相似的丰度分布，而基因簇 2 中的基因则表现出相反分布特征。列对应于按第一个共同主成分 (CPC 1) 排序的单个细胞，CPC 1 与 RNA 和蛋白质数据集中的细胞类型都有很强的相关性

(E) 基因簇 1 的 GO 分析

(F) 基因簇 2 的 GO 分析

(G) 对 Cluster 1 中的基因进行分子复合物检测

(H) cluster 2 中的基因 GO 分析得到的词条进行关联分析

1.4 使用 CITE-seq 的公共数据进行转录组和蛋白组的联合分析

在对质谱产生的数据进行以上分析后，虽然发现两种受到转录正向和负向调控的基因，但这一发现并不是由真正的单细胞多组学实验产生的，这种对角线整合模式的说服力不及垂直整合。

CITE-seq 是一种可以直接获得部分膜蛋白和转录组的测序方法，利用 CITE-seq 技术获取的公共数据集进行分析，以确认在细胞分群方面蛋白质组学相较于传统 RNA 测序是否提供了更丰富的信息。

仅使用 RNA 数据并应用 UMAP 算法进行降维分析，识别出 23 种不同的细胞类型。进一步，通过对免疫细胞标记物 CD45 进行特征绘图，我们可以清晰地区分出这些细胞群中的免疫细胞群体（图 8A）。

采用加权最近邻 (WNN) 算法结合细胞表面蛋白数据与转录组数据进行分析，观察到更为细分的细胞亚群（图 8B），表明膜蛋白数据提供了额外的分辨率，有助于揭示细胞群体中的微妙差异。

选择与免疫反应有关的巨噬细胞进行分析，将膜蛋白上用抗体标记的蛋白（都是 CD 蛋白）与对应的转录组数据相对齐，最终得到了 8 个基因供分析。将它们对细胞的转录组以及蛋白表达量相整合（与 1.3 中提到的方法一致），得到了这些基因受到转录后调控的相关性指标（图 8C）。这些基因都属于 cluster 1，且相关指数很高。对于这些基因进行 GO 分析，结果显示它们主要富集在生物正向过程、免疫过程和对外界免疫刺激的响应过程中，与之前的分析结果相符（图 8D）。这一系列分析有助于增强对基因调控和功能的理解，尤其是在单细胞多组学数据的基础上进行细致的研究。

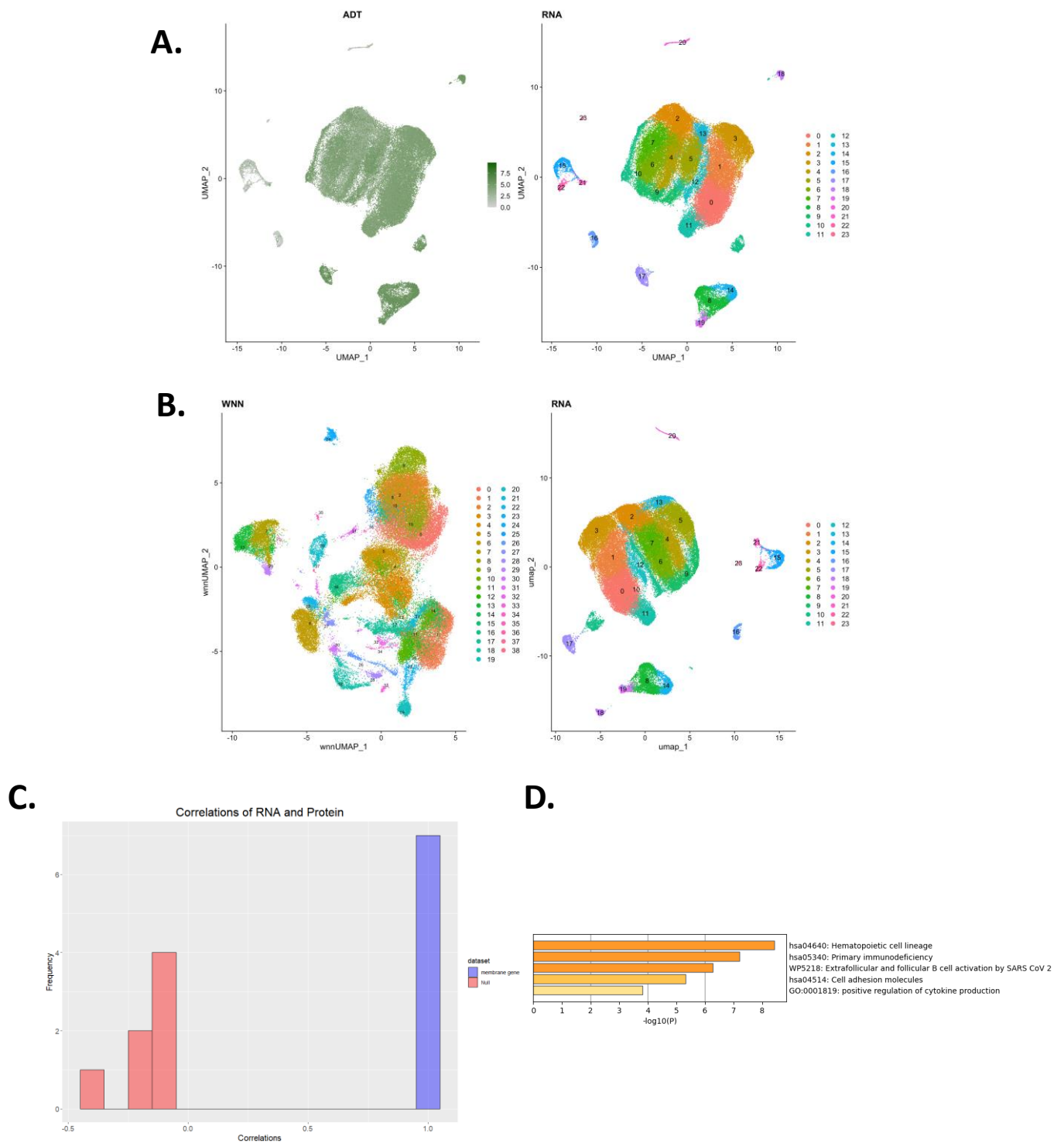


图 8：使用 CITE-seq 的公共数据进行转录组和蛋白组的联合分析

- A：仅使用转录组数据进行分群，用 CD45 Pro-Mask 查看免疫细胞分布
- B：使用 WNN 整合蛋白和转录组数据，分出更多亚群
- C：分析方法如图 3C，此图为零假设和所用的 8 个基因进行相关性计算的直方图
- D：对所分析的蛋白进行 GO 分析

2. 大脑蛋白组

取成年雄性小鼠大脑，除去小脑和嗅球后分离皮质海马，用 8M 尿素裂解后在梯度跑开观察实验结果，之后使用胶内酶解的方法提取蛋白，为尽可能覆盖

所有质核比段的蛋白，共上样 6 针，在皮质中鉴定到 7101 种蛋白，海马中鉴定到 6933 种蛋白（图 9A），后续也验证了一些小胶质细胞特异性蛋白如 S100B 等，在不同组间普遍存在。

不同脑区也有百种差异表达的蛋白（图 9B、C），这说明在大脑不同区域间细胞的异质性可以被蛋白组的差异所表征。

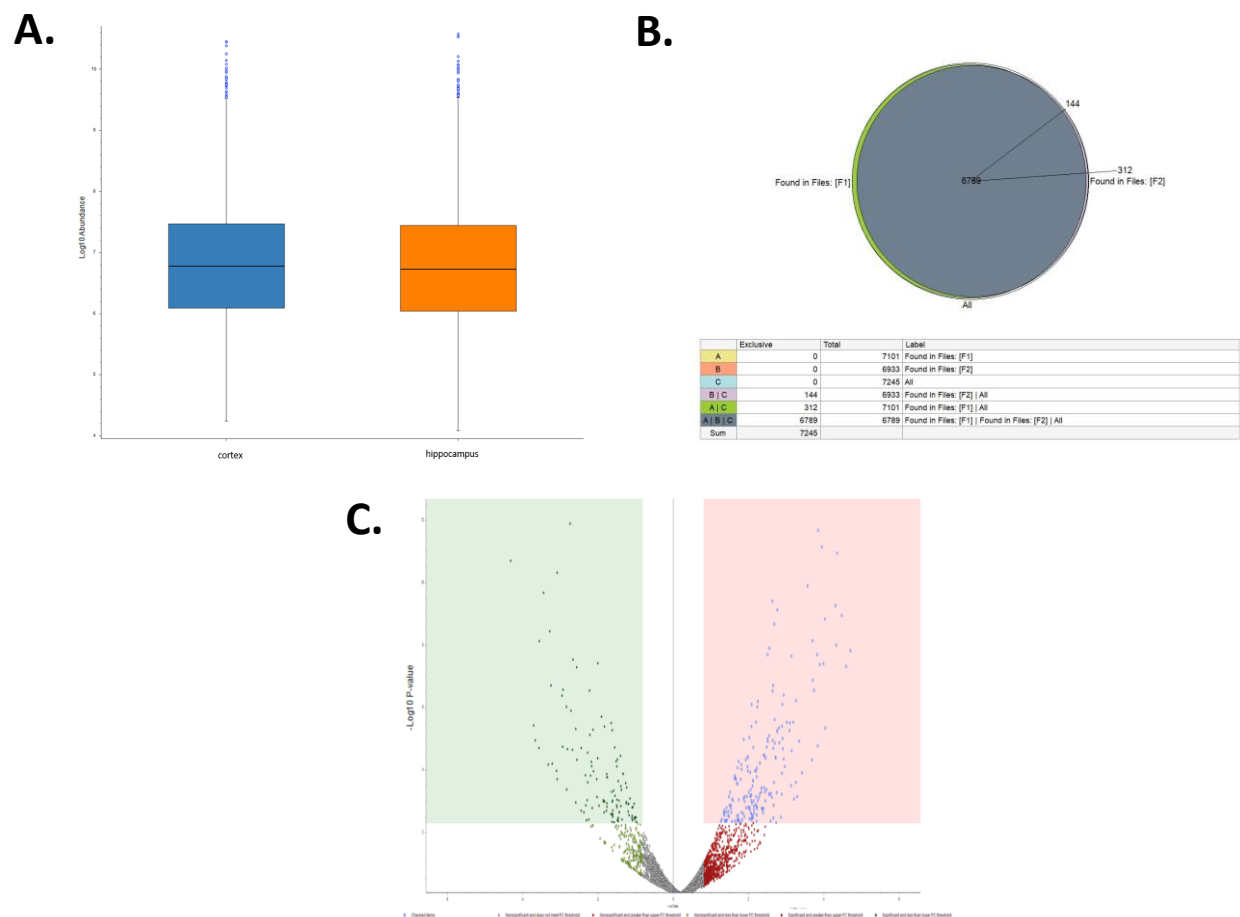


图 9：大脑蛋白组数据分析

- (A) 不同脑区鉴定到的蛋白数目
- (B) 各样本中蛋白量的韦恩图，F1 为皮质，F2 为海马
- (C) 不同脑区间蛋白的差异表达蛋白

3. 单细胞蛋白组实验

3.1 复现单细胞蛋白组实验

第一轮实验使用半年前准备好的 T 细胞 carrier 样本，经过单细胞样本的制备（裂解，TMT 标记，酶解等）和混样后上机，但本次实验的噪音信号（128N）甚至高于蛋白质的样本信号（126）（图 10）。

推测原因为实验中的 carrier 和 reference 的制备时间早，储存时间过长使得本就不多的蛋白质发生了降解，以及由于色谱柱出现问题，样本在-20° 冰箱中储存了一周才上样，该过程过长也可能造成了样品的降解。

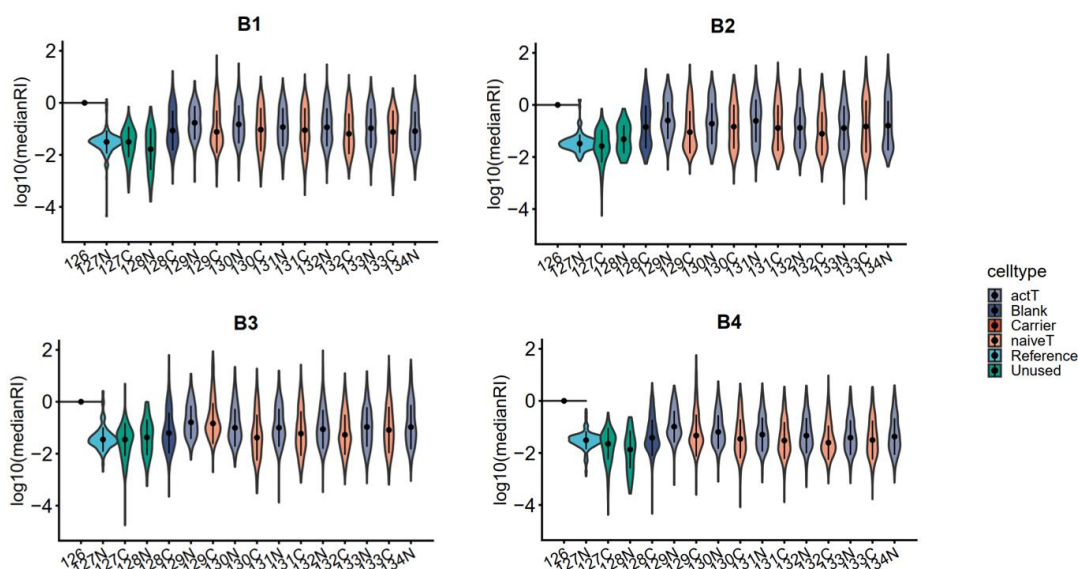


图 10：单细胞实验数据质量分析

后续从头开始养细胞的过程中，遇到了细菌污染和支原体污染问题，严重拖慢了实验进度。

3.2 单细胞蛋白组 carrier 蛋白释放效率优化

此实验全部为笔者设计，使用十二烷基- β -D-麦芽糖苷（DDM）和辛基葡萄糖苷（Octyl glucoside）两种裂解剂，TMT 作为内参定量分析裂解剂是否能提升蛋白的释放效率。由于质谱数据在期末季后半段才拿到，目前暂未对实验结果进行有效分析。不过从实验中对 SCoPE2 的反思已融合到之后的蛋白组实验设计规划中。

六、参考文献

1. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M. & Mann, M. 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* 11, O111.009613 (2012).
2. Slavov, N. Counting protein molecules for single-cell proteomics. *Cell* **185**, 232–234 (2022).
3. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
4. Doherty, M. K., Hammond, D. E., Clague, M. J., Gaskell, S. J. & Beynon, R. J. Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. *J. Proteome Res.* **8**, 104–112 (2009).

5. Ramazi, S. & Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database* **2021**, baab012 (2021).
6. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **55**, 182–196 (2017).
7. Tamara, S., Den Boer, M. A. & Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* **122**, 7269–7326 (2022).
8. Nguyen, H. P. & Schug, K. A. The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications. *J. Sep. Sci.* **31**, 1465–1480 (2008).
9. Shuken, S. R. An Introduction to Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **22**, 2151–2171 (2023).
10. Doerr, A. DIA mass spectrometry. *Nat. Methods* **12**, 35–35 (2015).
11. Kelly, R. T. Single-cell Proteomics: Progress and Prospects. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
12. Bandura, D. R. *et al.* Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
13. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
14. Chung, H. *et al.* Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat. Methods* **18**, 1204–1212 (2021).
15. Gerner, M. Y., Kastenmuller, W., Ifrim, I., Kabat, J. & Germain, R. N. Histo-Cytometry: A Method for Highly Multiplex Quantitative Tissue Imaging Analysis Applied to Dendritic Cell Subset Microanatomy in Lymph Nodes. *Immunity* **37**, 364–376 (2012).
16. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 44447 (2017).
17. Triana, S. *et al.* Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).
18. Chen, A. F. *et al.* NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).

19. Li, Z.-Y. *et al.* Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Anal. Chem.* **90**, 5430–5438 (2018).
20. Woo, J. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat. Commun.* **12**, 6246 (2021).
21. Brunner, A. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
22. Meier, F. *et al.* Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **14**, 5378–5387 (2015).
23. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
24. Huffman, R. G. *et al.* Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nat. Methods* **20**, 714–722 (2023).
25. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
26. Cost-effectiveness of radical cystectomy versus trimodality therapy for muscle invasive bladder cancer. | Journal of Clinical Oncology.
https://ascopubs.org/doi/abs/10.1200/JCO.2020.38.6_suppl.575.
27. Chen, A. T., Franks, A. & Slavov, N. DART-ID increases single-cell proteome coverage. 399121 Preprint at <https://doi.org/10.1101/399121> (2019).
28. Heil, L. R. *et al.* Evaluating the performance of the Astral mass analyzer for quantitative proteomics using data independent acquisition.
29. Lancaster, N. M. *et al.* Fast and Deep Phosphoproteome Analysis with the Orbitrap Astral Mass Spectrometer.
<http://biorxiv.org/lookup/doi/10.1101/2023.11.21.568149> (2023)
doi:10.1101/2023.11.21.568149.
30. Streit, W. J., Braak, H., Xue, Q.-S. & Bechmann, I. Dystrophic (senescent) rather than activated microglial cells are associated with tau pathology and likely precede neurodegeneration in Alzheimer’s disease. *Acta Neuropathol. (Berl.)* **118**, 475–485 (2009).
31. Perry, V. H., Nicoll, J. A. R. & Holmes, C. Microglia in neurodegenerative disease. *Nat. Rev. Neurol.* **6**, 193–201 (2010).

32. Norden, D. M., Trojanowski, P. J., Villanueva, E., Navarro, E. & Godbout, J. P. Sequential activation of microglia and astrocyte cytokine expression precedes increased iba-1 or GFAP immunoreactivity following systemic immune challenge. *Glia* **64**, 300–316 (2016).
33. Sierra, A., Gottfried-Blackmore, A. C., McEwen, B. S. & Bulloch, K. Microglia derived from aging mice exhibit an altered inflammatory profile. *Glia* **55**, 412–424 (2007).
34. Hickman, S. E., Allison, E. K. & El Khoury, J. Microglial Dysfunction and Defective β -Amyloid Clearance Pathways in Aging Alzheimer's Disease Mice. *J. Neurosci.* **28**, 8354–8360 (2008).
35. Mathys, H. *et al.* Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. *Cell Rep.* **21**, 366–380 (2017).