

AI 革命席卷蛋白质测序领域

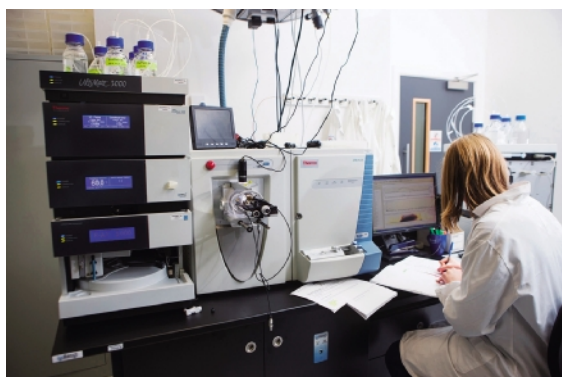
本报讯 蛋白质测序一直是医学诊断、环境和考古学研究面临的挑战。现在,人工智能(AI)正在改变这一领域——通过分析组成蛋白质的氨基酸序列从而识别蛋白质。与传统方法相比,AI 不仅识别速度更快,还能帮助研究人员对以前从未见过的蛋白质进行测序。

近日,一项发表于《自然-机器智能》的研究显示,名为 InstaNova 的 AI 可以识别伤口中的致病蛋白质和海水样本中微生物产生的未知蛋白质。

事实上,InstaNova 并非个例。在过去 4 年里,研究人员已开发出 20 多个蛋白质测序 AI 模型。“很明显,这是该领域的发展方向。”美国华盛顿大学蛋白质组学 AI 开发人员 William Noble 说。

蛋白质远比 DNA 和 RNA 复杂。人类基因组包含约 2 万个基因,但这些基因会产生 1000 万种不同的蛋白质。

生物学家通过将蛋白质分解成肽段来识别蛋白质,肽段由 5 到 20 个氨基酸组成。随后他们用质谱仪对这些肽段进行称重,并将其与数十个数据库中已知肽段的重量进行匹配,以确



研究人员希望用 AI 分析质谱仪数据,为复杂样本的蛋白质识别带来变革。图片来源:LEWIS HOUGHTON

定身份,最终再将这些肽段拼接成完整的蛋白质分子。

但这种方法存在局限性。比如,质谱法发现的多达 70% 的肽段并不存在于现有数据库中。

“传统蛋白质组学有点像谷歌搜索。如果它不在数据库中,你就找不到它。”丹麦技术大学的蛋白质组学专家 Timothy Patrick Jenkins 说,特别是随着肽数据库的不断扩充,计算机匹配所需的时间越来越长。

AI 则无须匹配肽段。它们计算了所有可能肽段的重量,而后者可能是由对一个给定长度的肽的化学修饰产生的。如果 AI 得出的片段与实际样本

中的肽段匹配,它就会尝试将其组装成完整的蛋白质。

为提高准确性,蛋白质测序 AI 接受了数百万个已知肽段及其如何组装成蛋白质的训练。这使得 AI 能够学习氨基酸链最常见的结合方式。

Jenkins 说,这种方法类似于大语言模型,就像 ChatGPT 通过大量文本训练学习语法规则一样。AI 也习得了一种蛋白质“语法”,为给定的一组肽提供最可能的序列。

2021 年,Noble 和同事推出了首个基于深度神经网络的蛋白质测序 AI——Casanovo。他们在 2024 年发表于《自然-通讯》的论文中报告说,该 AI 能有效识别训练数据中没有的新肽序列。实验表明,Casanovo 擅长识别免疫系统攻击癌症时靶向的细胞表面肽,以及海水样本中的未知蛋白质。

而 Jenkins 和同事开发的 InstaNova 在深度神经网络基础上引入了扩散策略。AlphaFold 等蛋白质结构预测模型

也采取了这种策略。

在与 Casanovo 的对比测试中,InstaNova 与升级版 InstaNova⁺ 结合,从实验室制造的 9 种生物的蛋白质混合物中成功识别出 42% 的肽段。

在真实蛋白质组学测试中,InstaNova 从感染的腿部伤口中鉴定出 1225 种血液白蛋白特有的肽,是传统方法检索结果的 10 倍。其中 254 种是数据库中不存在的新肽。

其他领域的研究人员也在使用蛋白质测序 AI。英国剑桥大学蛋白质组学研究员 Matthew Collins 最近就在测试几种蛋白质测序 AI 工具分析考古样本的能力。

Collins 指出,大多数情况下,考古样本的蛋白质由于在地下经过亿万年来发生了化学变化,或者它们来自早已灭绝的动植物,因此不太可能存在于传统蛋白质和肽数据库中,而这些 AI 模型尤其适用于混乱环境中的蛋白质检测。

利用 AI 工具,Collins 团队已在尼安德特人遗址中发现了兔子蛋白质特征,并在古代巴西的陶罐中检测到鱼类肌肉蛋白痕迹。(徐锐)

蛋白作祟让中年人患上痴呆

本报讯 痴呆通常会影响到老年人的生活,而当中年人出现相关症状时,往往难以识别。额颞叶痴呆(FTD)就是这样一种 60 岁以下人群中最常见的痴呆类型。美国研究人员发现了 FTD 发病机制的新线索,有望推动新型诊断方法的升级,并帮助更多患者尽早进入相关临床试验。近日,研究论文发表于《自然-衰老》。

多发于中年人群的 FTD 常常被误诊为抑郁症、早发型阿尔茨海默病、帕金森病或其他精神类疾病。患者平均需要 3.6 年才能得到准确诊断,目前尚无治愈方法,也缺乏能够延缓或阻止病情发展的治疗手段。因此,找到发病机制并提前诊断和干预非常重要。

研究团队测量了来自 116 名 FTD 患者脊髓液的 4000 多种蛋白质,并将它们与 39 名健康亲属的脊髓液进行比较。这 116 人均患有遗传性 FTD,使得研究人员能够在活着的确诊患者身上研究该疾病——这是非遗传性 FTD 病例无法做

到的,因为后者只能在死后确诊。

研究显示,FTD 患者中发生变化的蛋白质表明,这些患者存在 RNA 调控异常,RNA 调控是大脑基因正常表达所必需的,同时患者还存在影响大脑连接功能的缺陷。研究人员表示,这些蛋白质可能是 FTD 在中年发病时最早出现的特异性生物标志物。

“FTD 影响的是处于人生黄金时期的群体,并剥夺了他们的独立能力。”论文通讯作者、美国加利福尼亚大学旧金山分校的 Rowan Saloner 说,“但不像阿尔茨海默病等其他类型的痴呆,目前还没有一种确定的方法可以诊断就诊患者为 FTD。”

“如果我们能尽早识别出 FTD,或许可以利用发现的某些蛋白质作为标志物,引导患者获得正确的资源,比如让他们参与合适的治疗试验并获得精准治疗。”Saloner 补充说。(蒲雅杰)

相关论文信息:

<https://doi.org/10.1038/s43587-025-00878-2>

本报讯 瑞士日内瓦大学的 Denis Jabaudon 团队提出祖细胞代谢的区域差异决定了发育过程中大脑的生长。相关研究近日发表于《细胞》。

研究人员构建了小鼠大脑的单细胞分辨率出生时间图谱。该图谱显示,后脑的神经发生是短暂的,且局限于早期发育阶段,前脑的神经发生则通过减少脑室区祖细胞的消耗性分裂而在时间上持续延长。此外,该图谱还揭示了不同脑区直接与间接神经发生的具体模式。

通过单细胞 RNA 测序,研究人员识别出进化上保守的细胞周期程序和与代谢相关的分子通路。这些通路控制着各区域增殖的时间窗口。通过体内功能获得与功能缺失实验,研究人员鉴定出在新皮层晚期富集的线粒体蛋白 FAM210B 是关键调控因子。FAM210B 通过延长线粒体并增加乳酸生成,促进祖细胞的自我复制性分裂,最终导致其子代克隆的规模更大。(柯讯)

相关论文信息:

<https://doi.org/10.1016/j.cell.2025.04.003>

不同脑区祖细胞代谢差异
塑造大脑发育生长模式