

当蛋白质组遇上 AI, 加速的是什么

● 本报记者 温才妃

一个普通人做一次血液全蛋白质组的质谱检测,能获得哪些有意义或有趣的信息?谁会成为蛋白组领域的23andMe(DNA鉴定公司)?

去年7月,当人类蛋白质组98.5%的蛋白质结构被AlphaFold算法破译时,就有不少网友在知乎提出了这一问题。还有一些投资者干脆直接询问上市公司,是否有涉及人工智能(AI)在生命科学领域的发展计划和技术储备。

虽然蛋白质组学研究并非新鲜概念,但随着AI技术的突破,蛋白质组学相关应用开发与市场化正在快速发酵。

今年年初,人工智能蛋白质组诊断服务提供商西湖欧米(杭州)生物科技有限公司宣布完成数亿元Pre-A轮融资。他们开发的用于甲状腺结节诊断的临床实验室自建项目(LDT)产品,也问世在即。

蛋白质组学+AI,会不会是下一个市场焦点?

新算法重塑蛋白质组学

有人调侃,蛋白质组学概念之所以在二级市场遇冷,首要原因是人们并不清楚它究竟是什么。

但其实,它是开启精准医疗的钥匙。

电影《我不是药神》中的神药“格列卫”,其靶点就是一个融合蛋白,通过药物小分子抑制融合蛋白,最终控制慢性髓性白血病的发展。换句话说,谁更“懂”蛋白质,谁就能找到那把破解重大疾病、研发新药的钥匙。

然而,“传统的蛋白质组学分析技术和方法,并不完全适合用来研究蛋白质系统,所缺乏的是对蛋白质进行定量数据积累的过程,而且没有一个合适的算法。”西湖大学特聘研究员、西湖欧米创始人郭天南说。

AlphaFold的出现让科学界看到了曙光。

2021年7月,人工智能公司DeepMind和欧洲生物信息研究所合作,发布由AlphaFold预测的蛋白结构数据库,完成了人类蛋白质组98.5%的蛋白质结构预测。这被认为是本世纪最重要的科学突破之一。

随后,DeepMind在《自然》公布了



科研人员使用压力循环技术处理微量组织样品。
西湖欧米供图

AlphaFold2的源代码。AlphaFold2可在几分钟内破译一般蛋白质的三维结构,还能预测一个由2180个氨基酸相连的大蛋白质结构。

“在AlphaFold2之前,AI技术一直没有真正进入微观的生命世界。但实际上,AI技术之于微观生命世界不仅是一个必要条件,也是一个充分条件。”郭天南说。

如果没有AI技术,人类的理解力不可能弄清如此多的蛋白质动态,因为它实在太复杂了。“基因相对稳定,一个人一辈子做一次基因测序就够了,但蛋白质无时无刻不在发生变化,一场感冒都会改变很多细胞、蛋白质。而精准医学就是根据病人当前的情况,给他提供最合适的诊断和治疗方法。有了蛋白质组学+AI,精准医学将会提高到另一个层面。”郭天南说。

AI 助力催生精准医疗

十几年前,郭天南在武汉协和医院血液科学习、工作。他清楚地记得,一盒格列卫30多万元,科室把它买下来,一颗一颗卖给病人。后来,他远赴瑞士苏黎世联邦理工学院,师从蛋白质组学领域的开拓者之一Ruedi Aebersold。

2017年归国后,他把高通量质谱技术带到了西湖大学实验室。简而言之,这项技术就是给数以万计的蛋白质“测体重”,通过精确到小数点后30位,去辨别A君、B君。要知道在微观世界,蛋白质始终处于变化之中,尚无任何技术可以通过分子表面特征去识别。而高通量质谱技术则可以给蛋白质的运动“拍电影”,当药物进入细胞后,记录蛋白质变化所产生的大数据。

“质谱用于记录分子量,高通量则要求快,从中推断出它是什么样的蛋白质、

有多少量。”郭天南告诉记者。

精准医疗主要涉及诊断与治疗两方面,西湖欧米在这两方面均有布局。

甲状腺结节是人群中的高发疾病,有30%的甲状腺结节无法识别恶性、良性,很多病人在心理压力之下挨了不必要的一刀,而失去甲状腺的病人须终身服药。西湖欧米将蛋白质组结合AI技术开发出的LDT产品,能够让这30%的病人中的绝大部分避免挨刀。

郭天南还有一个“小目标”,就是通过研发成本降至1/10的LDT产品,让其成为物美价廉的诊断方法。

而更让科学家与市场兴奋的是,AI也赋予制药更多想象空间。与国内不同,国外制药公司非常重视开发新的药物靶点,目前已发现的药物靶点约有500个,而约40%的药物是以G蛋白偶联受体为靶点发现和设计的。

在AI制药中,高通量质谱技术是公认的实际使用中最有效的测量选择。在郭天南看来,至少可以在两方面入手:产生与制药相关的蛋白质数据,联合AI模型,建立比药物筛选更有效的新方法,促进新药研发;建立一个基于蛋白质、大数据和AI技术的新药物生产与质控流程,找到最佳药物生产方法。

目前,他正在同多个国际制药公司合作开发药物靶点,并配合医院、药厂开展相关制药工作。

想象空间可能更大

对比二级市场尚未“显山露水”,近年来,蛋白质组学在一级市场迎来了它的“小阳春”。

2020年,蛋白质工程服务商拜谱生物宣布完成千万级Pre-A轮融资;蛋白质组学技术服务及产品销售商中科新生命宣布完成2亿元A轮融资;蛋白质组

学技术开发及应用商景杰生物宣布完成5.3亿元的B轮融资。

郭天南所在的西湖欧米于1月13日宣布完成数亿元Pre-A轮融资。该轮融资由倚锋资本和高瓴创投共同领投,高榕资本、幂方资本和西湖科创创投跟投。去年西湖欧米已完成了天使轮和天使+轮融资。

如今,蛋白质组学方法被应用于肾癌、肝癌、结直肠癌、肺癌、胃癌等癌症的临床研究和诊治。科学家们纷纷表示,蛋白质组学驱动精准医学(PDPM)新时代已经到来。

高榕资本执行董事乐贝林告诉记者,用蛋白质做诊断、药物抗体开发以及蛋白原料的公司此前受到资本青睐,它们分别代表了蛋白质在终端及原料中间体的应用。与此同时,布局蛋白质产业上游端,比如全新发现的大数据端蛋白质组,近年来也备受关注。

“以往,蛋白质组在样本大小、类型以及单位样本所产生的数据量上有所限制。因此,我们更青睐在基础数据方法学上有创新和独到的技术,并能产生高质量数据的公司。”乐贝林说。

掌握基础方法学最核心的技术,可实现极微量样本蛋白质组的分析,拥有多样化的临床样本处理能力、独到的数据分析解读能力,以及医学背景出身的郭天南是公司很好的“翻译官”……这些都是西湖欧米被资本方看中的原因。

从人类基因组计划启动到最近十年,境内外涌现了一批上市公司,如华大基因、肿瘤NGS公司GRAIL、蛋白质组学公司SEER。“蛋白质组学理论上可以复制基因组在诊断上的路线,甚至走得更加深入。进一步看,蛋白质组学的应用从诊断迈向治疗,想象空间可能更大。”乐贝林说,这也是对郭天南在该阶段尝试AI制药的信心和期待所在。

能否步入深水区,发现有代表性的药物靶点,并把它做成全新的药物,是科学家孜孜以求的目标。如中国科学院院士、军事医学科学院院长贺福初在肝癌方向发现了很好的靶点,并在靶点上进行了药物筛选。“下一步,蛋白质组学的技术积累能否在药物的创新性研发上有所突破,我们拭目以待。”乐贝林说。