潜在位点难预测 人工智能来帮忙

作者：writer 来源：爱科学

本文原地址：https://www.iikx.com/news/progress/15329.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

潜在位点难预测 人工智能来帮忙。基因组包括两类遗传信息——DNA序列遗传信息及表观遗传学信息。后者涉及何时、何地以何种方式去应用遗传学信息，是作物农艺性状多样性的基础。

近年来，表观遗传学成为生命科学热点领域，检测技术和高通量测序的发展推动了作物表观组学的研究。然而，受到多种因素制约，仍有大量表观修饰位点没有得到发掘和研究。

近日，中国农业科学院生物技术研究所谷晓峰课题组、田健课题组和普莉课题组在《新植物学家》（New Phytologist）上发表了最新研究成果。他们构建了植物表观遗传修饰智能预测在线工具SMEP。

该项工作利用人工智能，深度学习植物DNA甲基化、RNA甲基化、组蛋白修饰等序列信息，系统实现了水稻、玉米等物种中表观修饰位点的预测，为作物功能基因组研究和智能设计育种提供工具和数据支撑。

真核生物基因转录的主要驱动力

谷晓峰告诉《中国科学报》，表观遗传修饰是指在基因的DNA序列没有发生改变的情况下，基因功能发生了可遗传的变化，并最终导致了表型的变化。这些可遗传的表观修饰包括DNA甲基化修饰、RNA甲基化修饰、组蛋白修饰、非编码RNA、染色质重塑等。

表观遗传修饰是调控真核生物基因转录、RNA代谢以及其他生物过程的主要驱动因素。谷晓峰说，植物、动物等真核生物细胞正常功能的维持是DNA序列遗传、表观遗传这两种信息互相作用、保持平衡的结果。如果这两种因素的任何一种表达失衡，都有可能导致不正常的基因表达，不能形成正常的生长发育过程。

表观遗传在调控基因表达、RNA稳定性、蛋白质翻译等方面起着关键作用，参与作物产量、品质、抗逆等多个重要的生物学过程。

普莉介绍，作为主要表观修饰类型之一，甲基化可以发生在DNA、RNA或组蛋白尾部序列上，且甲基化类型具有多样性，不同位置、不同类型的甲基化修饰对基因的表达具有激活或抑制作用。

目前，以表观遗传学理论和技术为中心的精准表观育种打开了培育农作物全新路径的可能性。普莉解释说，表观遗传修饰位点是指在DNA、RNA或者组蛋白序列上的甲基化、乙酰化等修饰位

点。

解析作物的品质产量、广适应性和耐逆性等重要农艺性状的调控网络，鉴定重要的表观修饰位点，尤其是新的未知的修饰位点，可以为高产优质的作物新品种提供表观遗传基因资源、修饰位点等理论支持和技术支持。谷晓峰说。

然而，由于表观遗传修饰在不同发育时期、不同生长环境等条件下具有动态可逆的调控特性，受到检测技术、实验成本、取材组织的限制，目前仍有大量重要的参与调控基因表达、生长发育等的表观修饰位点没有得到发掘和研究。因此，如何利用现有表观遗传组学数据深入探索潜在的表观修饰位点是表观遗传研究面临的关键问题。田健说。

人工智能生物技术成为关键

目前，科学家已经获得了一些作物的表观遗传组学大数据。该用什么方法解析这些数据呢？

谷晓峰团队将目光投向了人工智能等信息科学的创新和应用。人工智能驱动基因编辑、合成生物、全基因组选择等前沿生物技术快速发展，催生了人工智能和生物技术的深度融合。这种深度融合产生的人工智能生物技术（AIbiotechnology）成为生物农业领域的关键技术。

在此背景下，以深度学习为核心的人工智能技术在处理大数据（包括图像数据、多组学数据）方面具有显著优势。

考虑到人工智能在处理大数据方面的优势，以及表观遗传动态可逆的特点，应用人工智能的技术预测修饰位点可以弥补由于检测技术、实验成本、取材组织造成的表观修饰位点鉴定数量不足的问题。利用深度学习算法和模型优势，我们可以训练AI分析表观遗传多组学大数据，从而获得有重要应用价值的潜在表观修饰位点。谷晓峰说。

谷晓峰介绍，通过深度学习和训练，该团队基于卷积神经网络（CNN）方法，利用课题组前期绘制的粳稻日本晴的多种表观修饰图谱，构建得到了具有高准确度的智能预测模型（SMEP）。

人工智能首先通过收集与整理已发表的重要表观遗传组学数据和修饰位点，然后利用深度学习的算法从中进行特征序列提取、特征序列学习训练、构建智能预测模型，然后再对基因组序列进行扫描，发现潜在的表观修饰位点。田健解释道。

普莉介绍，为了验证该模型的预测是否成功，他们采用多个准确率评估指标来评估预测准确率；利用两个不同水稻品种（籼稻和粳稻）的表观数据进行相互验证；将预测结果与已发表的传统机器学习方法进行比较，进一步验证该方法的准确率；利用表观遗传检测手段，验证预测位点的表观修饰水平变化。

经过参数优化、交叉验证以及实验验证，SMEP模型在预测DNA甲基化、RNA甲基化和组蛋白修饰等表观遗传修饰位点中具有高可信度。

免费在线智能预测

这个模型从数据库中预测到了水稻、玉米等作物中大量潜在的表观修饰位点，这些修饰位点是常规技术尚未检测到的，参与了调控基因表达、环境胁迫响应等重要过程。谷晓峰说。

水稻、玉米等是重要的粮食作物，其产量和品质受到全球变暖和极端高温的严重威胁而显著降低。研究显示，平均气温每上升1摄氏度，水稻产量就可能下降10%，玉米产量下降5.8%以上。

谷晓峰说，应对挑战的关键是系统阐明作物响应高温胁迫的生物学基础，鉴定和优化响应高温胁迫的关键基因和调控位点，培育抵抗高温胁迫的新品种。

他们研究发现，水稻中DNA腺嘌呤甲基化能够影响基因表达，从而响应高温胁迫的环境信号。在热胁迫响应中，主要通过动态调节DNA腺嘌呤甲基化水平和影响热胁迫响应基因的表达两种方式发挥作用。

为了验证预测位点的准确性，他们利用水稻幼苗期经高温胁迫（45℃）处理36小时的材料，进行DNA腺嘌呤甲基化特异抗体实验。他们选取三个热响应转录调节基因作为候选基因，根据预测的潜在的表观修饰位点所在区域设计引物。

结果表明，预测位点和实验结果吻合度很高，显示了智能预测位点的准确性和实用性。普莉说，预测的表观遗传修饰位点可以为水稻、玉米的耐高温育种提供有用的基因资源和位点，加速培育耐高温的作物新品种。

基因组数据非常庞大，其中还有更多表观遗传修饰位点待发现。为了让更多科学家关注和参与相关研究，该团队构建了SMEP在线智能预测工具，可向用户免费提供检索表观遗传修饰位点和基因表达数据的可视化界面。

科研人员可以快速、准确地查询水稻、玉米等作物基因组任何区域或任何基因已知的和潜在的表观修饰位点，为作物功能基因组研究、重要基因克隆和功能分析、作物设计育种提供数据支撑。田健说，他们还分享了所有数据和代码，从事人工智能研究的工作者也可以应用这些代码和模型。（来源：中国科学报李晨）

相关论文信息：https://doi.org/10.1111/nph.17630

SMEP网址：http://www.elabcaas.cn/smep/index.html

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。
作者：谷晓峰等 来源：《新植物学家》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发