
科研人员提出基于深度学习的RNA多类型修饰解析算法

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/38118.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！

科研人员提出基于深度学习的RNA多类型修饰解析算法

。RNA修饰对RNA剪接加工、出核转运，以及RNA稳定性和翻译效率具有重要调控作用。然而，现有研究大多局限于单一修饰类型，依赖免疫沉淀、化学转换等检测方法，难以实现多种RNA修饰的同时检测。因此，如何在同一转录本上系统解析不同修饰的分布模式，探究修饰间的协同或竞争关系，并阐明它们如何共同调控RNA剪接加工等关键生物学过程，是该领域面临的重要挑战。

近期，中国科学院动物研究所研究团队，提出了基于纳米孔RNA直接测序技术与深度学习策略的RNA修饰图谱解析算法ORCA。该算法通过对RNA修饰位点的整体识别及基于RNA修饰位点数据库的迁移学习注释，实现了多种RNA修饰类型的系统鉴定，且该算法利用纳米孔测序的单分子关联信息，精准解析了RNA中临近修饰位点间的复杂互作关系，提示了RNA修饰与剪接加工过程之间的潜在协同调控模式。

研究团队通过信号多态性特征提取与域对抗学习策略，克服了RNA修饰检测对特定修饰类型训练集的依赖问题。同时，算法ORCA利用RNA修饰在转录本上的“不完全性”，提取了修饰与未修饰RNA分子在同个碱基上产生的信号多态性特征，并结合域对抗学习策略，使模型能够选择性学习不同修饰类型间的共性特征。基于该策略，该算法仅需使用6种体外转录合成的RNA修饰数据集进行训练，即可在mRNA和核糖体RNA中，对超15种修饰类型进行系统识别与化学计量比预测，拓展了现有DRS数据中可解析的RNA修饰类型范围。

在此基础上，研究团队建立了基于迁移学习的RNA修饰类型注释方法。该方法综合利用同类修饰位点间的电信号扰动相似性与序列基序偏好性，并结合大量RNA修饰数据资源，对识别到的修饰位点进行系统注释及标签迁移训练，从而实现多种已知RNA修饰类型的可靠注释。同时，研究团队将背景修饰位点的随机采样作为负样本训练，有效避免了对未知新型修饰类型的错误分类。在保证结果可靠性的同时，明显扩充了已知RNA修饰位点的数量。

研究团队进一步利用SGNex项目中多个人类细胞系的RNA直接测序数据，构建了跨细胞类型的RNA修饰全景图谱，发现了RNA修饰位点在转录本上呈现系统的成簇分布。同时，研究团队基于纳米孔测序独特的单分子支持信息，建立了可在单分子水平上，对邻近修饰位点间的协同或互斥进行系统鉴定的修饰簇互作识别模型。该模型发现，在转录本异构体特异的修饰位点附近，剪接调控因子及修饰相关RNA结合蛋白的结合位点显著富集。这一结果表明，RNA修饰与剪接加工

过程之间存在系统性的功能关联，为深入研究转录组表观修饰多样性及其复杂的调控机制提供了研究思路。

该研究通过纳米孔RNA直接测序技术与人工智能算法深度融合，实现了RNA修饰检测种类与数量的明显提升，揭示了RNA修饰位点的系统性邻近分布及其协同调控模式，为解码RNA修饰的动态互作及其在不同生物学背景下的调控机制提供了计算生物学工具。

相关研究成果发表在《自然-通讯》（Nature Communications）上。研究工作得到国家自然科学基金委员会、科学技术部、浙江省等的支持。

[论文链接](#)

基于深度学习的RNA修饰系统识别与注释模型

研究团队单位：动物研究所

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](http://iikx.com)转发