科研人员研发RNA转录本水平剪接调控预测模型

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/39948.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！

科研人员研发RNA转录本水平剪接调控预测模型

。RNA剪接调控是多样转录本产生的核心过程，其调控模式影响转录本的序列、结构及功能特征。长读长转录组测序技术为解析同一基因中不同RNA转录本的相对比例、识别各类组织和细胞类型的特异表达转录本全长，提供了技术支撑。但受限于样本获取难、测序成本高等因素，可用于复杂生理状态下转录本水平剪接分析数据仍不足。现有算法可预测个别样品中剪接位点强度，却难以实现转录本水平的预测，且对未知组织及细胞类型的扩展能力有限。

近日，中国科学院北京基因组研究所（国家生物信息中心）科研团队研发了可用于转录本水平RNA剪接预测的模型HELIX。该模型通过整合基因组序列与组织特异性的RNA结合蛋白表达矩阵，可高精度预测RNA剪接与转录本使用模式，为解析复杂生理条件下的可变剪接提供了高性能工具。

HELIX依托上下文依赖剪接调控机制构建，采用层次化深度学习子模型嵌套策略。这一策略基于DNA序列预测剪接位点及其基线强度，结合1499个RNA结合蛋白的表达特征，精准预测具体样本中的剪接调控水平，最后利用基于嵌入继承的长短期记忆网络，解析多个剪接位点间的依赖关系。评估显示，HELIX在高度调控位点及转录本相对比例预测方面较现有主流方法具有优势。

研究表明，HELIX可解析疾病研究中的异常RNA剪接及转录本使用特征。团队基于大规模结直肠癌队列，识别了肿瘤细胞中剪接失调与转录本表达异常现象，发现其与基因组突变、RNA结合蛋白表达异常及临床特征密切相关，为理解肿瘤发生机制及患者分层提供了分子特征依据。

团队进一步研发了HELIX的单细胞扩展版本scHELIX，其可预测不同细胞类型及肿瘤亚群中转录本差异使用模式，为解析肿瘤内部异质性提供新视角。结果显示，不同肿瘤亚克隆之间存在明显的RNA剪接与转录本表达差异特征，为肿瘤演化研究及靶点发现提供了新思路。

这项研究完善了关于组织特异性和疾病相关剪接机制的认识，为癌症分型、致病变异解释以及精准医学研究提供了方法学支撑。

相关研究成果发表在《自然-计算科学》（*Nature Computational Science*）上。研究工作得到国家自然科学基金委员会等的支持。

论文链接

HELIX模型架构及训练数据

研究团队单位：北京基因组研究所

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转发