



纳米孔测序检测RNA修饰的算法开发现状和应用前景展望

罗婷婷, 徐默萍, 王森, 陈发迎, 史偲君*

同济大学生命科学与技术学院, 附属同济医院, 上海 200065

* 联系人, E-mail: shij@tongji.edu.cn

收稿日期: 2025-11-18; 接受日期: 2026-02-24; 网络版发表日期: 2026-05-29

摘要 RNA分子上已知存在超过170种化学修饰, 它们广泛参与RNA代谢与功能调控. 与传统间接检测方法不同, 牛津纳米孔直接RNA测序技术能够直接读取RNA分子, 记录不同修饰引起的电流信号扰动和碱基识别错误, 并利用计算建模进行精准区分, 从而实现在单次实验中同时检测多种RNA修饰. 本文系统综述了基于该技术的RNA修饰检测算法开发现状, 重点介绍了针对 N^6 -甲基腺苷、假尿苷、5-甲基胞嘧啶等关键修饰的有监督机器学习与差异比较两大类方法, 分析了其原理、特性与适用场景. 同时, 本文概述了该技术在全转录组修饰图谱绘制、修饰功能解析及在疾病生物标志物发现中的应用潜力, 并对其在多修饰同步检测、模型泛化能力及标准化分析方面面临的挑战与未来发展进行了展望.

关键词 RNA修饰, 纳米孔测序, 生物信息学算法, 表观转录组学

与仅含有十余种主要表观遗传修饰的DNA相比, RNA上已知存在超过170种不同的共价修饰^[1], 统称为“表观转录组”^[2]. RNA修饰广泛参与转录、剪切^[3]、翻译^[4]以及介导蛋白质结合^[5]等调控过程, 深刻影响RNA的代谢命运和细胞稳态. N^6 -甲基腺苷(N^6 -methyladenosine, m^6A)是真核生物信使RNA(messenger RNA, mRNA)中含量最丰富的修饰, 主要由METTL3, METTL14和WTAP组成的甲基转移酶复合物催化形成^[6]. 研究表明, m^6A 既可促进mRNA降解从而调控其稳定性^[7], 又能提高mRNA的翻译效率^[8], 同时还参与调控前体mRNA的剪接, 特别是促进特定外显子的跳跃^[9]. 假尿苷(pseudouridine, Ψ)是丰度仅次于 m^6A 的另一常见RNA修饰, 可通过RNA依赖性或非依赖性机制

生成. 研究发现, Ψ 同样参与基因表达的精细调控, 例如促进mRNA剪接并有效提升其翻译效率^[10]. 除影响RNA代谢外, RNA修饰还可改变其物理性质, 例如由TRMT6-TRMT61A复合物介导的 N^1 -甲基腺苷(N^1 -methyladenosine, m^1A)修饰^[11], 其引入的正电荷会干扰碱基配对, 进而影响RNA的二级结构^[12].

为解析多样化的RNA修饰, 研究人员近年来开发了多种检测方法. 2012年, 首个可在全转录组范围内检测 m^6A 的方法——甲基化RNA免疫沉淀测序(methylated RNA immunoprecipitation sequencing, MeRIP-seq)问世^[13,14], 该方法结合特异性抗体免疫沉淀与二代测序(next-generation sequencing, NGS), 为绘制RNA修饰图谱开辟了新途径. 除MeRIP-seq这类基于

引用格式: 罗婷婷, 徐默萍, 王森, 等. 纳米孔测序检测RNA修饰的算法开发现状和应用前景展望. 中国科学: 生命科学, 2026, 56: 1893–1907
Luo T T, Xu M P, Wang M, et al. Nanopore sequencing for RNA modification detection: current algorithm development and future applications (in Chinese). Sci Sin Vitae, 2026, 56: 1893–1907, doi: 10.1360/SSV-2025-0377

抗体识别修饰核苷酸的方法^[13,14]外, 还有一类方法依赖修饰特异性化学反应富集带有特定修饰的读段, 例如检测m⁶A的GLORI、检测Ψ的BID-seq等^[15,16]。基于NGS的RNA修饰检测方法推动了该领域的快速发展, 但由于其无法直接读取RNA修饰, 必须依赖抗体或化学诱导等间接手段进行推断, 因此通常一次实验仅能检测一种修饰, 难以同步绘制多种修饰图谱, 且操作流程繁琐, 易受批次效应干扰^[17,18]。

三代测序的快速发展为高效绘制RNA修饰图谱带来了新的机遇。牛津纳米孔技术(Oxford Nanopore Technologies, ONT)推出的直接RNA测序(direct RNA sequencing, DRS)方案无需抗体或化学标记, 可直接对天然RNA分子进行测序。自2017年问世以来, DRS技术已历经多次版本升级(从SQK-RNA002至SQK-RNA004), 在孔道稳定性、测序通量及碱基识别准确性方面均取得显著提升。该技术可在单核苷酸分辨率下捕获两类关键特征(图1)^[19]: 一是RNA分子通过膜嵌入式纳米孔蛋白时引起的离子电流扰动; 二是RNA修饰引发的碱基识别错误, 例如Ψ常导致U→C错配^[20], 5-甲基胞嘧啶修饰则易引起下游插入或缺失事件^[21]。不同修饰核苷酸产生的信号特征存在差异, 可通过计算建模进行区分, 因此DRS原则上具备检测任意RNA修饰的潜力^[22]。迄今为止, DRS已成功应用于检测m⁶A、N⁷-甲基鸟苷(N⁷-methylguanosine, m⁷G)、5-甲基胞嘧啶(5-methylcytosine, m⁵C)、5-羟甲基胞嘧啶(5-hydroxymethylcytosine, hm⁵C)、Ψ、N⁴-乙酰胞苷(N⁴-acetylcytidine, ac⁴C)和2'-O-甲基化(2'-O-methylation, Nm)等多种RNA修饰类型^[17,20,23,24]。

为从DRS数据中检测和定量不同RNA修饰, 近年来已涌现出大量计算方法(图2)。本文系统回顾了基于纳米孔测序的RNA修饰检测算法与工具, 详细总结了各方法的算法原理、适用修饰类型及其技术特点(表1)。此外, 本文还综述了基于DRS的RNA修饰检测在机制研究及疾病诊断中的应用, 并对该领域面临的发展瓶颈与未来机遇进行了总结与展望。

1 纳米孔RNA测序的技术原理与文库构建

纳米孔RNA测序技术的实现, 依赖于两个核心环节: 一是将RNA分子制备成适合测序的文库, 二是基于电信号检测的核苷酸测序。当施加跨膜电压时, 带

负电的核酸分子被驱动穿过嵌于膜上的蛋白质纳米孔, 不同核苷酸通过时会引起特征性电流变化。这些实时的电流变化被高灵敏度传感器记录, 并通过机器学习算法解码为对应的核苷酸序列, 从而实现无需逆转录或扩增的直接测序。

完整的RNA分子需要经过特异性建库才能被有效测序, ONT为直接测序天然RNA分子开发了专用的文库制备方案。该方案的核心步骤是将测序接头连接到RNA的3'末端, 有两种主要方式: 一是将接头直接连接至天然RNA的3'端, 无需常规逆转录步骤; 二是先合成一条互补DNA(complementary DNA, cDNA)链, 形成RNA-cDNA杂交双链, 再进行接头连接。在这两种策略中, 均只有RNA链穿过纳米孔。前者操作步骤更少、速度更快, 能最大程度保持RNA的天然状态, 适用于快速检测; 后者形成的文库更稳定, 支持更长的测序读长, 因此通常产出更高。直接RNA测序平均准确度通常较低, 约为83%~86%^[25,26]。因此, 研究者在选择建库策略时, 需在保持RNA修饰信息、操作简便性与速度(直接连接法)以及文库稳定性与产出(杂交链法)之间, 根据具体研究目标进行权衡^[27]。

2 m⁶A修饰检测计算方法

2.1 有监督的机器学习方法

有监督的机器学习方法是当前纳米孔测序中RNA修饰检测的主流策略之一。其核心思想是从已知修饰状态的训练数据中学习信号特征(如电流强度、驻留时间等)与特定RNA修饰之间的映射关系, 进而对未知样本中的修饰位点进行预测。这类方法通常依赖参考数据集的构建, 例如利用m⁶A-seq^[14]、Ψ-seq^[15]或体外合成修饰转录本等提供模型训练所需的真实标签。

根据训练数据来源, 有监督学习方法可分为两类: (i) 基于真实生物学样本训练的模型; (ii) 基于体外人工合成(*in vitro* transcription, IVT)RNA训练的模型。前者能更好地反映生理条件下的信号特征, 但训练数据中常存在样本异质性与噪声; 后者可获得更理想化、均质的训练样本, 但可能低估真实样本中的复杂性(如不同修饰间的信号串扰)。

在算法层面, 有监督方法多采用机器学习或深度学习框架, 包括随机森林(random forest, RF)、支持向

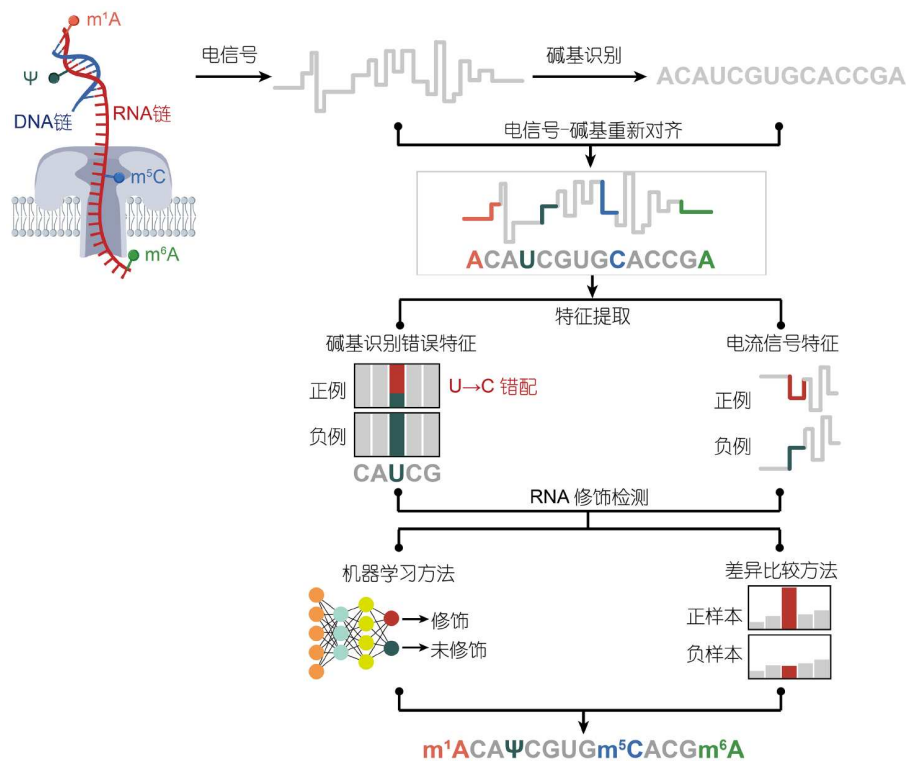


图 1 基于DRS的RNA修饰检测原理与策略. 部分元素使用BioGDP绘制^[19]

Figure 1 Principle and strategy of RNA modification detection based on DRS. Some elements were created using BioGDP^[19]

量机(support vector machine, SVM)、卷积神经网络(convolutional neural network, CNN)、循环神经网络(recurrent neural network, RNN)等. 这些算法能够捕捉信号的局部模式与上下文依赖性, 对每个碱基或 k -mer的修饰状态进行分类或概率预测. 部分模型还引入多示例学习(multi-instance learning, MIL)或迁移学习(transfer learning)策略, 以在不同修饰类型或物种间共享特征表示, 从而提升模型的适用性.

2.1.1 基于真实生物学样本训练的模型

基于真实生物学样本训练的模型使用来自真实细胞环境的转录组数据, 并结合抗体免疫沉淀(如miCLIP, m6ACE-seq)或化学标记(如GLORI, DART-seq)等高精度检测技术确定的修饰位点作为真实标签, 使模型能更准确地捕捉天然状态下的修饰信号.

早期研究主要依赖传统机器学习算法实现位点水平的 m^6A 预测. 其中, MINES^[28]利用随机森林模型, 基于HEK293细胞系的miCLIP^[29]数据训练, 通过提取纳米孔信号中的电流强度、停留时间及碱基上下文等特

征进行修饰识别. 这类方法特征可解释性强, 但在复杂信号建模和跨样本泛化方面存在局限.

深度学习的引入显著提升了模型对信号模式的提取能力. DENA^[30]基于双向长短期记忆网络(bidirectional long short-term memory, Bi-LSTM), 利用拟南芥野生型与 m^6A 缺失突变体的DRS数据, 学习修饰与未修饰信号的动态差异. m6Abasecaller^[31]基于深度神经网络, 使用HEK293细胞系的修饰位点数据训练, 将碱基识别与修饰预测融为一体, 实现同步分析. m6Aiso^[32]采用残差网络结构和半监督学习对DART-seq内源性标记的单分子 m^6A 电信号进行训练, 实现单分子水平的 m^6A 检测. NanoSPA^[33]作为一个多功能模型, 结合不同神经网络模块实现两种修饰的识别: m^6A 修饰预测采用前馈神经网络, 以HeLa细胞的m6A-SAC-seq数据作为训练集; Ψ 修饰预测则采用扩展卷积网络, 以混合来源的核糖体RNA(ribosomal RNA, rRNA) Ψ 位点数据训练.

为解决信号覆盖不均与标注不确定性带来的挑战, 基于MIL的框架应运而生, 该类模型能够同步实现

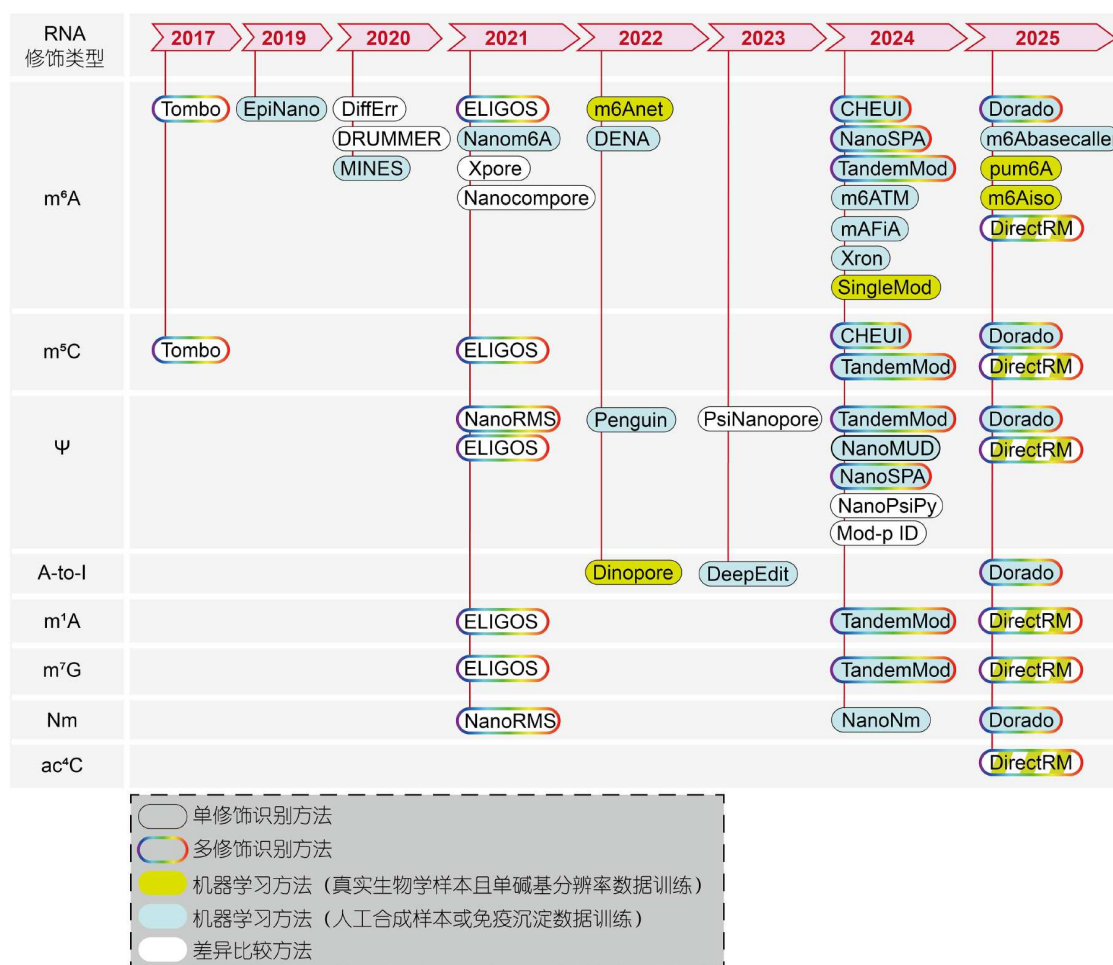


图 2 基于DRS的RNA修饰检测算法发展情况。黑色边框为仅支持单一修饰的检测算法；彩色边框为支持多种修饰类型的检测算法，如Tombo支持m⁶A和m⁵C的检测。绿色背景为基于真实生物学样本的单碱基分辨率数据(如GLORI)训练的深度学习算法；蓝色背景为基于体外人工合成RNA(IVT)或非单碱基分辨率数据(如免疫共沉淀MeRIP-seq)训练的深度学习算法；白色背景为差异比较类检测算法。DirectRM的训练数据是HepAD38细胞系的MeRIP-seq富集峰与数据库中其他样本来源修饰位点的交集，作者未在单碱基水平直接证实这些位点是否在HepAD38中真实存在，因此用绿色条纹标注

Figure 2 Evolution of algorithms for RNA modification detection based on DRS. Black borders indicate algorithms that support single modification type; colored borders indicate algorithms that support multiple modification types (e.g., Tombo supports the detection of both m⁶A and m⁵C). Green background denotes deep learning algorithms trained on single-base resolution data from real biological samples (e.g., GLORI); blue background denotes deep learning algorithms trained on *in vitro* transcribed (IVT) RNA or non-single-base resolution data (e.g., immunoprecipitation-based MeRIP-seq); white background denotes differential analysis algorithms. The training data for DirectRM consists of the intersection of MeRIP-seq peaks from HepAD38 and modification sites from other samples in public databases. Since the authors did not validate at single-base resolution whether these sites are genuinely present in HepAD38, DirectRM is marked with a green striped background

单分子与位点层面的高分辨率预测。m6Anet^[34]率先将MIL框架应用于RNA修饰检测领域^[35]，利用HEK293T细胞的m6ACE-seq数据训练，并兼容RNA002与RNA004两个DRS版本。pum6A^[36]在m6Anet的基础上进行优化，融合了碱基配对特征与原始信号，其训练数据整合了HEK293T的miCLIP，m6ACE-seq与GLORI数据集。SingleMod^[37]以不同细胞系和物种的GLORI

数据作为训练集，包括人类(HEK293T, HeLa)、植物(拟南芥, 水稻)和绿藻(莱茵衣藻)，同样兼容不同试剂盒版本；该模型采用“每个基序一个模型”的设计策略，不仅增强了模型对特定序列上下文的识别特异性，还在修饰水平定量预测方面表现出优越性能。

近期，多标签学习框架的引入为同时检测多种修饰提供了新思路。与传统的“单修饰单模型”策略不同，

表 1 基于纳米孔测序的RNA修饰检测算法特性概览

Table 1 Overview of characteristics of RNA modification detection algorithms based on nanopore sequencing

检测工具	适用RNA修饰类型	特征类型	算法原理	是否可定量	单分子水平 预测	文献来源
Tombo	m ⁶ A, m ⁵ C	电流信号	Fisher检验	是	是	[46]
EpiNano	m ⁶ A	碱基错配	支持向量机	否	否	[17]
MINES	m ⁶ A	电流信号	随机森林	否	否	[28]
DiffErr	m ⁶ A	碱基错配	G检验	否	否	[44]
DRUMMER	m ⁶ A	碱基错配	G检验	否	否	[45]
ELIGOS	m ⁶ A, m ¹ A, 5moU, Ψ, m ⁷ G, hm ⁵ C, m ⁵ C, f ⁵ C	碱基错配	Fisher精确检验	否	否	[47]
Nanom6A	m ⁶ A	电流信号	极端梯度提升	是	是	[43]
NanoRMS	Ψ, Nm	碱基错配+电流信号	K最近邻域法	否	否	[21]
xPore	m ⁶ A	电流信号	高斯混合模型	是	是	[49]
Nanocompore	m ⁶ A	电流信号	高斯混合模型	否	否	[48]
DENA	m ⁶ A	电流信号	双向长短期记忆网络	是	是	[30]
Dinopore	A-to-I	碱基错配+电流信号	卷积神经网络	否	否	[55]
m6Anet	m ⁶ A	电流信号	多示例学习	是	是	[34]
PsiNanopore	Ψ	碱基错配	自定义的统计检验	否	否	[53]
NanoSPA	m ⁶ A, Ψ	碱基错配	极端随机树(Ψ), 前馈神经网络(m ⁶ A)	否	否	[33]
CHEUI	m ⁶ A, m ⁵ C	电流信号	卷积神经网络	是	是	[39]
TandemMod	m ⁶ A, m ⁵ C, hm ⁵ C, Ψ, m ¹ A, m ⁷ G	电流信号	双向长短期记忆网络	是	是	[23]
NanoMUD	Ψ, m ¹ Ψ	电流信号	双向长短期记忆网络	是	是	[50]
SingleMod	m ⁶ A	电流信号	深度神经网络+多示例学习	是	是	[37]
pum6A	m ⁶ A	碱基错配+电流信号	基于正例和无标签数据的多示例学习	是	是	[36]
Xron	m ⁶ A	电流信号	卷积循环神经网络+非齐次隐马尔可夫模型	是	是	[41]
Dorado	m ⁶ A, m ⁵ C, Ψ, A-to-I, Nm	电流信号	未披露	是	是	https://software-docs.nanoporetech.com/dorado/
m6Aiso	m ⁶ A	电流信号	残差网络	是	是	[32]
m6ATM	m ⁶ A	碱基错配+电流信号	扩张卷积神经网络+双流多示例学习	是	是	[42]
mAFiA	m ⁶ A	电流信号	逻辑回归	否	否	[40]
NanoNM	Nm	电流信号	极端梯度提升, 随机森林	否	否	[24]
DirectRM	m ⁶ A, m ⁵ C, Ψ, m ¹ A, m ⁷ G, ac ⁴ C	碱基错配+电流信号	双向长短期记忆网络	是	是	[38]
m6Abasecaller	m ⁶ A	电流信号	深度神经网络	否	否	[31]
Penguin	Ψ	碱基错配+电流信号	支持向量机, 随机森林, 神经网络	否	否	[51]
NanoPsiPy	Ψ	碱基错配	卡方检验	否	否	[54]
Mod-p ID	Ψ	电流信号	Fisher检验	否	否	[52]
DeepEdit	A-to-I	电流信号	卷积神经网络	否	否	[56]

多标签学习使单个模型能够同时识别和区分不同的RNA修饰类型。DirectRM^[38]作为一个集成的多标签模型, 采用两阶段学习策略: 首先利用二分类模型筛选包含候选修饰的 k -mers, 随后通过基于注意力的神经网络精确定位修饰位点。该模型的训练数据是HepAD38细胞系的MeRIP-seq富集峰与数据库中其他样本来源修饰位点的交集, 综合运用原始电流信号与碱基识别错误作为输入特征, 实现了对天然RNA中六种修饰(ac^4C , m^1A , m^5C , m^7G , m^6A , Ψ)的同时检测, 代表了当前多修饰集成分析的最新进展。

2.1.2 基于体外人工合成RNA训练的模型

基于体外人工合成RNA训练的模型为RNA修饰检测提供了另一有效途径。EpiNano^[17]作为早期代表, 利用支持向量机算法, 通过分析体外合成的完全修饰与完全未修饰RNA样本的特征差异, 实现 m^6A 修饰的推断, 为后续研究奠定了基础。TandemMod^[23]采用Bi-LSTM架构, 从包含 m^6A , m^5C 和 m^1A 等多种修饰的体外合成RNA样本及其未修饰对照中学习序列上下文与信号的时间依赖性, 展现了其同时检测多种修饰类型并跨平台泛化的能力。CHEUI^[39]利用卷积神经网络对原始信号片段进行建模, 通过体外合成的 m^6A 与 m^5C 完全修饰样本训练, 支持单分子水平的修饰概率输出, 是一种高通量且可扩展的检测框架。mAFiA^[40]采用逻辑回归模型, 对来自体外合成RNA的电流信号、停留时间特征进行统计建模, 构建了一个线性且可解释的预测模型。Xron^[41]基于卷积递归神经网络框架, 能够从原始信号中提取特征, 兼容RNA002与RNA004试剂盒, 同时支持单分子水平与位点水平的概率预测。m6ATM^[42]结合Wavenet卷积结构与双重多示例学习框架, 有效捕获信号的局部细节与全局依赖; 该模型同样使用体外RNA训练, 并支持单分子与位点水平的预测。Nanom6A^[43]则采用极端梯度提升(extreme gradient boosting, XGBoost)算法, 以原始电流信号为输入特征利用体外合成RNA数据训练, 在毛果杨和哺乳动物数据上进行了验证。

此外, ONT公司官方推出的Dorado(<https://software-docs.nanoporetech.com/dorado/>)模型集成了多种RNA修饰检测功能。它通过统一的深度学习框架, 在完成标准碱基识别的同时, 直接输出包括 m^6A , m^5C , Ψ 和A-to-I编辑在内的多种修饰预测, 兼容RNA002与

RNA004两种测序试剂盒, 是目前计算效率最高的集成化检测方案之一。

2.2 差异比较的分析方法

差异比较分析是另一种常见的RNA修饰检测策略。该方法通常依赖于实验样本与对照样本在信号特征或碱基识别错误率上的差异, 通过统计检验或概率建模识别存在显著差异的位点, 从而推断潜在的RNA修饰。根据算法设计, 这些方法可分为基于显著性检验的统计方法与基于生成模型的贝叶斯方法两类。

在统计检验类方法中, 典型代表包括DiffErr^[44], DRUMMER^[45], Tombo^[46]和ELIGOS^[47]等。DiffErr与DRUMMER均采用G检验评估不同条件下的碱基识别错误率差异。DiffErr通过比较目标样本与参考样本在特定位点的错误碱基分布变化, 筛选出具有显著差异的位点, 并进一步结合重复样本的内部比较以提高预测可靠性。DRUMMER采用类似策略, 将修饰与无修饰样本之间的碱基识别比例差异作为输入, 并对显著差异位点进行多重检验校正, 从而识别可能的 m^6A 修饰。与之不同, ELIGOS以体外转录数据为基础, 通过比较修饰与未修饰RNA的碱基识别误差模式, 利用Fisher精确检验每个位点的误差比例差异, 从而检测多种RNA修饰类型。Tombo是最早基于DRS检测RNA修饰的开源工具, 该软件集成了从序列比对到信号分配的全套分析流程, 通过将测序信号重新映射至参考序列, 实现碱基事件的信号层级解析。Tombo提供三种修饰检测模式: (i) 基于已训练信号模型的检测; (ii) 不依赖模型的*de novo*检测; (iii) 基于差异比较的检测。其中, 差异比较模式以体外合成无修饰样本作为对照, 通过统计检验方法分析实验组与对照组在每个 k -mer上信号分布的差异, 从而识别潜在修饰位点。

与统计检验法不同, 贝叶斯建模方法通过对纳米孔信号的概率分布进行建模, 能够直接在读段水平上推断修饰概率, 并定量估计修饰比例。该类方法的代表包括Nanocompore^[48]和xPore^[49], Nanocompore通过提取每个碱基事件的电流信号强度及停留时间, 使用高斯混合模型(Gaussian mixture model, GMM)比较实验组与对照组信号分布的差异, 从而识别潜在修饰位点。xPore进一步扩展了这一框架, 引入多样本贝叶斯GMM模型, 对跨样本的信号均值进行聚类分析, 同时

区分修饰与未修饰RNA, 并估计每个位点的修饰水平.

有监督的机器学习方法能够在无需配对对照样本的情况下对未知样本进行直接预测. 此类方法在通用性、自动化程度和检测效率上表现突出, 尤其适用于大规模转录组修饰图谱绘制和跨样本比较分析. 差异比较类方法通常依赖实验设计中的无修饰对照样本, 但受限于实验条件、测序噪声及对照样本获取成本, 实际应用范围相对有限.

3 Ψ 修饰检测计算方法

3.1 有监督的机器学习方法

由于已发表的真实生物样本 Ψ 图谱仍然有限, 大多数 Ψ 检测工具依赖于体外人工合成RNA数据构建训练集. 例如, TandemMod作为一个通用型工具, 利用Bi-LSTM从包含 Ψ 的体外合成RNA样本中学习特征, 实现对包括 Ψ 在内的多种修饰的同时检测. NanoMUD^[50]则针对 Ψ 和 $m^1\Psi$ 开发, 同样基于体外合成RNA数据训练Bi-LSTM模型以实现精准预测. ONT官方推出的Dorado也集成了基于体外合成RNA数据训练的 Ψ 检测模型, 提供了便捷的解决方案. 相比之下, 基于真实样本数据训练的 Ψ 模型较少, NanoSPA是其中一个代表, 其训练数据来自HeLa细胞系的已知 Ψ 位点; 还有Penguin^[51]提供多种机器学习模型(如支持向量机、神经网络和随机森林)来识别 Ψ , 使用HEK293和HeLa数据集进行训练. 近期提出的多标签模型DirectRM同样基于真实样本数据训练, 并支持 Ψ 的检测, 反映了该修饰在表观转录组中的重要性以及相关检测技术的逐步成熟.

3.2 差异比较的分析方法

差异比较的分析方法无需模型训练, 而是通过统计手段比较实验组与无修饰对照(通常为体外合成RNA)的信号差异来推断 Ψ 位点. NanoRMS^[21]采用无监督的 k -means聚类算法, Mod-p ID^[52]则基于Fisher精确检验, 这两种方法均通过分析电流信号分布的差异实现修饰位点检测. 而PsiNanopore^[53]基于Fisher精确检验, NanoPsiPy^[54]则采用卡方检验, 二者均通过评估特定位点U $\rightarrow$ C错误率的显著性. 这些方法虽计算简单、结果易于解释, 但其性能高度依赖于对照样本的质量和测序深度.

4 m^5C 修饰检测计算方法

在有监督的机器学习方法中, TandemMod在体外合成RNA数据上训练, 能够捕获 m^5C 独特的信号特征. CHEUI使用卷积神经网络对体外合成RNA数据的原始信号进行建模, 能够输出单分子水平的修饰概率. DirectRM同样具备 m^5C 检测能力, 其优势在于基于真实生物样本数据训练. 官方模型Dorado也支持 m^5C 检测, 提供了高度集成的解决方案.

在差异比较类方法中, Tombo及其扩展模块Tombo_com是早期代表. 它们通过信号重映射与统计检验比较实验组与对照组信号分布, 以识别潜在修饰位点. 总体而言, m^5C 检测方法正逐步从统计检验过渡到深度学习建模, 以实现更高的跨样本泛化性能.

5 其他类型修饰的计算检测方法

5.1 m^1A 与 m^7G 修饰的检测

m^1A 与 m^7G 均属于甲基化修饰, 分别影响RNA翻译起始与5'端稳定性. TandemMod模型在训练中使用体外合成RNA数据, 通过Bi-LSTM网络学习不同修饰间的差异, 实现对 m^1A 与 m^7G 的联合检测. 其联合建模策略为探索不同甲基化修饰间的电信号特征差异提供了统一思路. DirectRM基于真实生物样本数据训练, 利用其注意力机制, 能够在天然RNA样本中实现对这两种修饰的特异性识别.

5.2 A-to-I编辑的检测

RNA的A-to-I编辑由ADAR酶家族催化产生, 常发生在非编码RNA和双链结构区域. Dinopore^[55]是专为A-to-I检测设计的模型, 结合碱基错配特征与电信号特征, 采用CNN实现特征提取与修饰预测. 该模型使用H9细胞系的训练数据进行建模, 能够有效区分腺苷与肌苷的信号差异, 是目前较为系统的A-to-I识别框架. DeepEdit^[56]是一个全连接神经网络模型, 其训练数据来源于在裂殖酵母中引入人类RNA编辑酶构建的A-to-I编辑菌株, 以及不具有RNA编辑的野生型菌株. 模型通过比较这两类菌株的纳米孔电流信号差异进行训练, 从而实现A-to-I编辑事件的识别. 此外, ONT官方识别工具Dorado也集成支持A-to-I编辑检测功能, 为该领域提供了新的分析方案.

5.3 Nm修饰的检测

Nm修饰广泛存在于rRNA、转运RNA(transfer RNA, tRNA)和mRNA中, 在RNA稳定性与剪接调控中发挥作用. NanoNm^[24]采用XGBoost算法, 以体外合成RNA及天然rRNA中的Nm位点为训练数据, 通过电流强度、事件持续时间等信号特征, 实现对Nm修饰的预测. 该方法在不同RNA类型间表现出较好的跨样本泛化性. 此外, NanoRMS也具备Nm修饰位点的检测能力, 而Dorado现已将Nm修饰检测纳入其功能, 进一步丰富了该修饰的分析工具选择.

6 基于纳米孔测序的RNA修饰检测在基础研究和临床诊断中的应用

6.1 在生物学基础研究中的应用

传统RNA修饰检测方法只能检测一种RNA修饰类型, 且目前仅限于检测有限几种RNA修饰. 此外, 传统方法通常依赖间接信号, 如基于抗体的方法受限于抗体特异性与批次效应, 难以准确定量修饰化学计量; 化学或酶学辅助方法虽可提升分辨率, 却面临反应条件严苛、流程复杂且仅适用于特定修饰等局限^[57]. 更重要的是, 这些方法普遍需将RNA反转录为cDNA, 并通过PCR扩增, 不仅可能丢失原始修饰信息, 也难以在单分子层面直接关联修饰与转录本结构^[58]. 相比之下, DRS技术直接对天然RNA分子进行测序, 从根本上避免了反转录与扩增引入的系统偏差. 并可以在一次实验中检测多种RNA修饰(包括m⁶A, m⁵C, Ψ等), 为绘制表观转录组全景图谱提供了高效技术途径. 此外, DRS建库流程日益简化, 更有利于开展大规模样本队列或临床转化研究.

在多种生物体系中的研究已充分展示了纳米孔测序在修饰图谱绘制中的应用潜力. Workman等人^[59]对人类转录组进行DRS测序, 首次在单分子水平同时检测m⁶A与A-to-I编辑, 证明了纳米孔测序在多修饰联合识别方面的可行性. Li等人^[60]对秀丽隐杆线虫的转录组进行纳米孔测序, 发现编码区内的RNA修饰水平普遍高于非编码区, 且所有非U碱基的修饰概率在终止密码子下游均显著下降, 提示U可能在翻译终止中发挥特异性作用. Zhang等人^[61]通过对拟南芥转录组进行测序, 证明m⁵C修饰在幼苗和花蕾的移动mRNA中高度富集, 并进一步揭示m⁵C在促进mRNA跨组织长

距离运输中的功能. Jiang等人^[62]利用纳米孔直接RNA测序系统性描绘了小鼠在进食与禁食状态下多个器官的RNA修饰与转录图谱, 揭示了能量代谢状态与组织特异性转录后修饰之间的动态关系. 同时, 他们在人源化小鼠模型中^[63], 通过直接RNA测序刻画了人源肝脏组织中的代谢转录特征与修饰分布, 为跨物种肝脏代谢研究提供了高分辨率的表观转录组参考图谱. 纳米孔测序还被应用于解析病毒RNA修饰组, 为理解病毒感染与宿主互作提供了新的分子视角. Price等人^[45]结合DRS与RNA甲基化免疫共沉淀测序, 分析了腺病毒感染过程中宿主与病毒转录本的m⁶A修饰, 发现病毒感染晚期RNA的表达依赖于宿主甲基转移酶METTL3, 并证实m⁶A修饰对腺病毒RNA的有效剪接至关重要.

上述研究拓展了纳米孔测序在多组织、多状态、多物种背景下的应用范围, 表明其不仅可用于基础表观转录组图谱的绘制, 也能够揭示修饰与代谢状态、组织分化及疾病进展之间的关联规律, 为RNA修饰组学的系统性研究奠定了重要基础.

6.2 在疾病研究与诊断中的应用

与传统组织活检相比, 液体活检通过分析血液、唾液、尿液等体液样本实现疾病诊断与监测, 具有非侵入性、可重复采集等优势^[64]. 该领域研究主要聚焦于识别和检测循环生物标志物, 包括DNA、RNA、蛋白质和代谢物, 这些物质可能存在于循环肿瘤细胞^[65]、外泌体^[64]或游离形式, 如细胞游离DNA(cell free DNA, cfDNA)^[66]或RNA(cell free RNA, cfRNA)^[67]. 液体活检在癌症诊断中的早期商业化应用主要集中于循环肿瘤DNA(circulating tumor DNA, ctDNA), 例如GRAIL公司基于cfDNA甲基化开发的诊断产品已在泛癌筛查中显示出良好前景^[68]. 尽管以质谱为代表的传统方法已能鉴定临床样本中的RNA修饰^[69,70], 但基于测序, 尤其是能在单次实验中同时捕获序列与多种修饰信息的DRS, 在临床应用中仍处于起步阶段.

DRS技术能够从单个RNA分子直接捕获RNA修饰信息, 从而捕捉可能来自目标组织和/或细胞(如肿瘤细胞)的独特RNA修饰图谱. 这一特性使其在发现和验证新型RNA修饰标志物方面展现出巨大潜力. 近年来, DRS检测RNA修饰技术的疾病筛查潜力已在多种体系中得到初步验证. 在细胞系^[48,53]和模式生物^[71,72]

中的研究验证了其技术可行性; 在智力发育障碍携带 *METTL5* 突变的患者血液样本^[73]中的初步应用展示了其检测潜力; 尤其在肺癌患者组织活检样本^[74]中, DRS已成功识别出新型癌症特征, 这些探索为未来将RNA修饰用于患者分层和辅助诊断提供了理论支持。

在疾病诊断应用中, 不同RNA类型通过DRS技术展现出其作为生物标志物的独特价值。mRNA长期以来是该领域的研究核心, 其差异表达分析已被广泛用于疾病分型与预后评估。近年来, 研究焦点逐渐扩展到mRNA本身的修饰变化, 尤其是在疾病状态下的修饰改变^[75], 特别是癌症领域^[76]。尽管基于DRS的mRNA研究仍处于早期阶段, 但已有先驱工作成功利用该技术绘制了人外周血样本中的全转录组与表观转录组图谱^[73], 为将mRNA及其修饰用于液体活检诊断奠定了方法学基础。

相较于mRNA, rRNA作为细胞中最丰富的RNA, 其极高的修饰程度^[77~79]和动态变化特性^[80]正成为新兴的生物标志物来源。一项关键研究表明, 在配对的肺癌与正常组织队列中, 仅凭rRNA修饰谱即可准确区分癌变与正常样本^[74], 这有力证明了rRNA修饰在疾病分类, 特别是肿瘤诊断中具有巨大应用潜力。

另一类可通过DRS检测的极具潜力的生物标志物是tRNA, 它们兼具高丰度与高度修饰的特性。研究显示, tRNA在癌症等多种疾病中普遍存在表达水平与修饰的失调^[81~83], 因而被视为潜在的新型诊断标志物与治疗靶点^[84], 然而, 基于NGS的tRNA研究方法存在明显局限, 其建库流程复杂^[85], 且对RNA修饰的检测能力有限。相比之下, DRS技术展现出显著优势。近期研究表明, 该技术可在单次实验中同步获取tRNA表达量与修饰信息^[86]。随着DRS测序试剂盒的更新(如RNA004), 已实现对tRNA分子的高通量覆盖, 并成功识别出43种不同RNA修饰的特征信号, 甚至能够检测低丰度的线粒体tRNA与病毒来源tRNA^[87]。这些发现在概念上证明了特定RNA修饰谱与疾病状态存在强关联, 凸显了DRS技术在推动RNA修饰从基础研究走向临床诊断应用中的关键价值。

然而, 要将DRS检测RNA修饰转化为具备明确临床灵敏度与特异度的诊断工具, 仍面临一系列严峻挑战。液体活检样本中RNA含量极低与DRS建库所需高起始量之间存在根本矛盾, 尽管样本多重测序可作为潜在解决方案, 但目前尚缺乏成熟稳定的实施方案。

循环RNA(如细胞游离RNA和外泌体RNA)多以短片段形式存在^[88]。虽然近期更新的算法使DRS已能检测到50 nt左右的读长, 但对于更短的、富含生物标志物信息的RNA, 其捕获效率仍然有限。而技术平台本身的快速迭代——试剂盒化学版本的频繁更新伴随着信号特征的变化, 导致下游生物信息学工具(包括修饰检测算法)迅速过时, 严重阻碍了标准化流程的建立和结果的长期可比性, 以及临床检测所需的可重复性。此外, 从原始信号到碱基识别的一致性仍需更多验证, 模型内部随机性(如初始化权重)和灵敏度不足(如靠近阈值)可能导致相同数据在多次分析中产生不一致的结果, 这不符合临床检测对可重复性的严格要求。这些因素共同制约着DRS技术在临床场景中的可靠应用。

尽管存在这些挑战, 但清晰的改进路径正在形成。通过开发超低起始量的建库技术、推动化学与生物信息学流程的标准化、建立权威的基准数据集, 并在大规模前瞻性队列中严格验证标志物性能, DRS技术有望从一项强大的基础研究工具逐步转化为可靠的临床诊断平台。当前的研究正在为这一转化奠定关键的生物学与方法学基础, 而克服上述瓶颈将是实现其临床潜力的决定性步骤。

7 总结与展望

纳米孔直接RNA测序是RNA修饰检测技术的一项重大突破, 为表观转录组学研究开启了全新视角。与依赖反转录、化学诱导或免疫富集的间接检测方法相比, DRS能够在单分子、单碱基分辨率下读取RNA分子的天然序列及修饰信号, 从而在一次实验中同时捕获序列与不同类型修饰信息, 为系统性描绘RNA修饰图谱、解析修饰动态调控以及揭示其生物学功能提供了强有力的技术支撑。

然而, 目前基于纳米孔的修饰检测方法仍面临若干挑战。如前所述, 差异比较类算法通过比较实验样本与配对的未修饰对照样本来识别RNA修饰位点, 虽然受实验噪声的影响较小, 但需要额外的实验投入; 机器学习类算法无需对照样本, 成本较低、使用便捷, 但性能易受训练数据质量的影响。此外, 大多数现有模型仅能检测单一或少数修饰类型, 而实际转录组中修饰种类复杂且常存在共存和互作现象^[39], 例如, 同一tRNA分子上可同时存在十余种不同修饰, 不同修饰之

表 2 RNA修饰检测代表性工具核心性能指标汇总

Table 2 Summary of core performance metrics of representative tools for RNA modification detection

修饰类型	检测工具	AUROC	召回率 ^{a)}	F1值 ^{a)}	预测位点与真实位点转录组分布相似度 ^{b)}	不同修饰水平预测稳健性 ^{c)}	运行时间(h)	兼容测序试剂盒版本
m ⁶ A	CHEUI	0.72	0.49	0.53	0.60	0.87	58.91	RNA002
	ELIGOS	0.63	1.00	0.37	0.67	0.74	2.63	RNA002
	m6Anet	0.78	0.64	0.55	0.63	0.71	6.66	RNA004/ RNA002
	Nanom6A	0.71	0.66	0.31	0.69	0.84	10.45	RNA002
	NanoSPA	0.71	0.55	0.58	0.92	0.69	18.98	RNA002
	Tombo	0.60	0.52	0.31	0.62	0.86	4.87	RNA002
	MINES	0.47	0.89	0.24	0.61	1.00	6.54	RNA002
	Epinano	0.72	0.52	0.18	0.75	0.70	1.76	RNA002
	DENA	0.66	0.73	0.21	0.88	0.94	9.27	RNA002
	SingleMod	0.90	0.84	0.39	0.77	0.81	4.36	RNA004/ RNA002
	TandemMod	0.73	0.58	0.29	0.89	0.75	12.36	RNA002
	pum6A	0.76	0.38	0.33	0.78	0.01	7.41	RNA002
	m6Aiso	0.64	1.00	0.14	0.88	1.00	44.36	RNA002
	Differr	0.56	1.00	0.25	0.70	0.50	5.98	RNA002
	ELIGOS_diff	0.65	0.74	0.39	0.72	0.62	83.93	RNA002
	Tombo_Com	0.72	0.51	0.35	0.72	0.75	5.94	RNA002
	DRUMMER	0.51	1.00	0.30	0.74	0.57	162.00	RNA002
	xPore	0.54	1.00	0.23	0.70	0.91	6.41	RNA002
	Epinano_delta	0.34	1.00	0.23	0.74	1.00	1.63	RNA002
	Nanocompore	0.65	0.19	0.29	0.92	0.01	41.85	RNA002
Ψ	Xron	0.57	0.83	0.52	0.82	0.63	15.23	RNA004/ RNA002
	Dorado	0.90	0.81	0.36	0.79	0.11	6.45	RNA004
	NanoRMS	0.72	0.14	0.20	0.60	0.84	14.30	RNA002
	PsiNanopore	0.85	0.42	0.38	0.72	0.99	96.56	RNA002
	NanoMUD	0.60	0.67	0.00	0.62	1.00	2.43	RNA002
	TandemMod	0.76	0.87	0.00	0.89	0.99	45.29	RNA002
	NanoSPA	0.66	0.03	0.05	0.80	0.50	16.59	RNA002
m ⁵ C	Dorado	0.50	0.12	0.01	0.60	0.02	8.94	RNA004
	Tombo	0.47	0.36	0.01	0.70	0.92	5.92	RNA002
	Tombo_Com	0.52	0.89	0.01	0.80	0.98	6.02	RNA002
	CHEUI	0.56	0.07	0.01	0.54	0.82	50.13	RNA002
	TandemMod	0.46	0.97	0.00	0.92	0.99	47.30	RNA002
A-to-I	Dorado	0.62	0.45	0.00	0.80	0.75	12.39	RNA004
	Dinopore	0.74	0.05	0.05	0.43	0.01	23.10	RNA002
m ¹ A	Dorado	0.68	0.48	0.00	0.35	0.64	9.18	RNA004
m ¹ A	TandemMod	0.43	0.93	0.00	0.90	1.00	48.62	RNA002
m ⁷ G	TandemMod	0.35	1.00	0.00	0.59	1.00	41.56	RNA002

a) 计算召回率和F1值时均使用达到最大F1值时的分类阈值. b) 相似度定义为1减去KS检验统计量(D). c) 稳健性定义为1减去不同修饰水平下相对召回率的标准化方差

间存在协同或拮抗作用^[89]. 理想的检测算法应能在单一平台上实现多修饰并行检测, 但目前距离“全修饰检测”的目标尚有较大差距.

纳米孔测序平台目前仍处于快速迭代阶段, 近期DRS试剂盒从RNA002更新至RNA004版本, 在提升准确率和通量的同时, 也导致电信号特征的变化, 使得针对RNA002的算法无法直接用于RNA004数据的分析. 利用统一的性能评估标准与涵盖多种修饰类型的基准数据集, 前人对已有算法的性能进行了跨样本、跨DRS版本的系统比较^[90-92]. 结果显示, 虽然m⁶A检测工具整体性能较好, 但目前的算法对m⁶A以外的修饰检测效力仍然不足, 生物学合理性和跨样本泛化能力欠佳, 并且容易对信号相似的修饰产生误判. 这些系统评估工作全面展现了各算法的优缺点, 为未来开发更完善的模型提供了宝贵的指导意见.

工具选择的复杂性进一步加剧了分析挑战, 用户需从近年涌现的多种检测工具中作出选择, 而不同工具在参数设置、检测偏好与性能表现上差异显著. 即便针对同一DRS数据集, 不同工具所识别的修饰位点也可能大相径庭, 其精确度与召回率往往依赖于用户设定的阈值. 为帮助研究者有效应对这一复杂性, 我们建议一些实用原则: 首先, 确保所选工具与当前测序化学版本(RNA002或RNA004)兼容, 此举可排除多数过时方法; 其次, 除目标样本外, 实验设计中如果包含无修饰对照样本, 则可以考虑使用差异比较类工具, 否则应使用机器学习类工具; 再次, 在分析中引入已知修饰状态的基准数据集对工具及参数进行验证, 例如利用公共可获取的完全修饰与未修饰序列进行性能评估, 以准确控制假阳性.

前人在真实生物学样本来源的独立测试集, 对多

种RNA修饰检测工具的十余项指标进行了全面比较^[92], 表2收录了其部分指标. 根据其测评结果, 研究者可考虑以下工具: 在m⁶A修饰检测方面, 推荐使用SingleMod, 其准确性不输于官方工具Dorado, 并在区分修饰酶敲除组和对照组的任务中表现更好. 对于非m⁶A修饰, 目前的工具都难以兼顾召回率和精确率. 对于Ψ, 如实验设计中包含无修饰对照样本, 则可以使用差异比较类工具PsiNanopore, 否则推荐Tandem-Mod, 它在机器学习类工具中展现出较为出色的受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)与召回率. 对于m⁵C修饰, 推荐官方工具Dorado, 其准确性和计算效率略优于其他机器学习工具. 对于A-to-I, 推荐Dinopore, 其准确性优于Dorado. 而对于m¹A和m⁷G等修饰, 可用模型较少, 尚待领域的进一步发展. 该工作除比较已有工具外, 还对部分工具进行了跨修饰类型和跨试剂盒版本的重新训练, 发现可以进一步提升其性能, 这为缓解工具短缺的现状提供了新思路. 另外, 若需使用多种同类工具并简化分析流程, 研究者可采用集成式标准化 workflow, 如MasterOfPores^[93]或Meta-Compore^[48].

总之, 开发能够融合碱基识别与多修饰检测功能的综合性分析工具, 是推动纳米孔RNA测序技术发展的关键路径. 纳米孔直接RNA测序技术有望成为系统绘制全转录组修饰图谱、解析RNA修饰动态调控规律的核心平台. 随着测序通量提升、实验成本优化及算法生态日趋完善, 这项技术将在精准医学、发育生物学及疾病机制研究等领域展现出更广阔的应用前景, 最终推动RNA修饰研究迈向全面化、量化与临床转化的新阶段.

参考文献

- 1 Boccaletto P, Baginski B. MODOMICS: an operational guide to the use of the RNA modification pathways database. *Methods Mol Biol*, 2021, 2284: 481–505
- 2 Saletore Y, Meyer K, Korlach J, et al. The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol*, 2012, 13: 175
- 3 Boo S H, Kim Y K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med*, 2020, 52: 400–408
- 4 Hernandez-Alias X, Benisty H, Schaefer M H, et al. Translational efficiency across healthy and tumor tissues is proliferation-related. *Mol Syst Biol*, 2020, 16: e9275
- 5 Lewis C J T, Pan T, Kalsotra A. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat Rev Mol Cell Biol*, 2017, 18: 202–210

- 6 Wang X, Huang J, Zou T, et al. Human m⁶A writers: two subunits, 2 roles. *RNA Biol*, 2017, 14: 300–304
- 7 Wang X, Lu Z, Gomez A, et al. N⁶-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 2014, 505: 117–120
- 8 Meyer K D, Patil D P, Zhou J, et al. 5' UTR m⁶A promotes cap-independent translation. *Cell*, 2015, 163: 999–1010
- 9 Adhikari S, Xiao W, Zhao Y L, et al. m⁶A: signaling for mRNA splicing. *RNA Biol*, 2016, 13: 756–759
- 10 Karijovich J, Yi C, Yu Y T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nat Rev Mol Cell Biol*, 2015, 16: 581–585
- 11 Saffra M, Sas-Chen A, Nir R, et al. The m¹A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature*, 2017, 551: 251–255
- 12 Li X, Xiong X, Wang K, et al. Transcriptome-wide mapping reveals reversible and dynamic N¹-methyladenosine methylome. *Nat Chem Biol*, 2016, 12: 311–316
- 13 Meyer K D, Saletore Y, Zumbo P, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 2012, 149: 1635–1646
- 14 Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature*, 2012, 485: 201–206
- 15 Carlile T M, Rojas-Duran M F, Zinshteyn B, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, 2014, 515: 143–146
- 16 Pandolfini L, Barbieri I, Bannister A J, et al. METTL1 promotes let-7 microRNA processing via m7G methylation. *Mol Cell*, 2019, 74: 1278–1290.e9
- 17 Liu H, Begik O, Lucas M C, et al. Accurate detection of m⁶A RNA modifications in native RNA sequences. *Nat Commun*, 2019, 10: 4079
- 18 Jora M, Lobue P A, Ross R L, et al. Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. *Biochim Biophys Acta*, 2019, 1862: 280–290
- 19 Jiang S, Li H, Zhang L, et al. Generic Diagramming Platform (GDP): a comprehensive database of high-quality biomedical graphics. *Nucleic Acids Res*, 2025, 53: D1670–D1676
- 20 Smith A M, Jain M, Mulroney L, et al. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One*, 2019, 14: e0216709
- 21 Begik O, Lucas M C, Prysacz L P, et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol*, 2021, 39: 1278–1291
- 22 Garalde D R, Snell E A, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*, 2018, 15: 201–206
- 23 Wu Y, Shao W, Yan M, et al. Transfer learning enables identification of multiple types of RNA modifications using nanopore direct RNA sequencing. *Nat Commun*, 2024, 15: 4049
- 24 Li Y, Yi Y, Gao X, et al. 2'-O-methylation at internal sites on mRNA promotes mRNA stability. *Mol Cell*, 2024, 84: 2320–2336.e6
- 25 Keller M W, Rambo-Martin B L, Wilson M M, et al. Direct RNA sequencing of the coding complete influenza A virus genome. *Sci Rep*, 2018, 8: 14408
- 26 Pitt M E, Nguyen S H, Duarte T P S, et al. Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA nanopore sequencing. *GigaScience*, 2020, 9: giaa002
- 27 Chen Y, Davidson N M, Wan Y K, et al. A systematic benchmark of nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. *Nat Methods*, 2025, 22: 801–812
- 28 Lorenz D A, Sathe S, Einstein J M, et al. Direct RNA sequencing enables m⁶A detection in endogenous transcript isoforms at base-specific resolution. *RNA*, 2020, 26: 19–28
- 29 Linder B, Grozhik A V, Orlarier-George A O, et al. Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat Methods*, 2015, 12: 767–772
- 30 Qin H, Ou L, Gao J, et al. DENA: training an authentic neural network model using nanopore sequencing data of *Arabidopsis* transcripts for detection and quantification of N⁶-methyladenosine on RNA. *Genome Biol*, 2022, 23: 25
- 31 Cruciani S, Delgado-Tejedor A, Prysacz L P, et al. *De novo* basecalling of RNA modifications at single molecule and nucleotide resolution. *Genome Biol*, 2025, 26: 38
- 32 Guo W, Ren Z, Huang X, et al. Single-molecule m⁶A detection empowered by endogenous labeling unveils complexities across RNA isoforms. *Mol Cell*, 2025, 85: 1233–1246.e7

- 33 Huang S, Wylder A C, Pan T. Simultaneous nanopore profiling of mRNA m⁶A and pseudouridine reveals translation coordination. *Nat Biotechnol*, 2024, 42: 1831–1835
- 34 Hendra C, Pratanwanich P N, Wan Y K, et al. Detection of m⁶A from direct RNA sequencing using a multiple instance learning framework. *Nat Methods*, 2022, 19: 1590–1598
- 35 Koh C W Q, Goh Y T, Goh W S S. Atlas of quantitative single-base-resolution N⁶-methyl-adenine methylomes. *Nat Commun*, 2019, 10: 5636
- 36 Liu C, Liang H, Wan A H, et al. Decoding the m⁶A epitranscriptomic landscape for biotechnological applications using a direct RNA sequencing approach. *Nat Commun*, 2025, 16: 798
- 37 Xie Y Y, Zhong Z D, Chen H X, et al. Single-molecule direct RNA sequencing reveals the shaping of epitranscriptome across multiple species. *Nat Commun*, 2025, 16: 5119
- 38 Zhang Y, Wu Y, Ma J, et al. DirectRM: integrated detection of landscape and crosstalk between multiple RNA modifications using direct RNA sequencing. *Nat Commun*, 2025, 16: 9450
- 39 Acera Mateos P, J Sethi A, Ravindran A, et al. Prediction of m⁶A and m⁵C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. *Nat Commun*, 2024, 15: 3899
- 40 Chan A, Naarmann-de Vries I S, Scheitl C P M, et al. Detecting m⁶A at single-molecular resolution via direct RNA sequencing and realistic training data. *Nat Commun*, 2024, 15: 3323
- 41 Teng H, Stoiber M, Bar-Joseph Z, et al. Detecting m⁶A RNA modification from nanopore sequencing using a semisupervised learning framework. *Genome Res*, 2024, 34: 1987–1999
- 42 Yu B, Nagae G, Midorikawa Y, et al. m6ATM: a deep learning framework for demystifying the m⁶A epitranscriptome with nanopore long-read RNA-seq data. *Brief Bioinf*, 2024, 25: bbae529
- 43 Gao Y, Liu X, Wu B, et al. Quantitative profiling of N⁶-methyladenosine at single-base resolution in stem-differentiating xylem of *Populus trichocarpa* using nanopore direct RNA sequencing. *Genome Biol*, 2021, 22: 22
- 44 Parker M T, Knop K, Sherwood A V, et al. Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m⁶A modification. *eLife*, 2020, 9: e49658
- 45 Price A M, Hayer K E, McIntyre A B R, et al. Direct RNA sequencing reveals m⁶A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun*, 2020, 11: 6016
- 46 Stoiber M, Quick J, Egan R, et al. *De novo* identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv*, 2016, 094672
- 47 Jenjaroenpun P, Wongsurawat T, Wadley T D, et al. Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res*, 2021, 49: e7
- 48 Leger A, Amaral P P, Pandolfini L, et al. RNA modifications detection by comparative nanopore direct RNA sequencing. *Nat Commun*, 2021, 12: 7198
- 49 Pratanwanich P N, Yao F, Chen Y, et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol*, 2021, 39: 1394–1402
- 50 Zhang Y, Yan H, Wei Z, et al. NanoMUD: profiling of pseudouridine and N¹-methylpseudouridine using Oxford Nanopore direct RNA sequencing. *Int J Biol Macromol*, 2024, 270: 132433
- 51 Hassan D, Acevedo D, Daulatabad S V, et al. Penguin: a tool for predicting pseudouridine sites in direct RNA nanopore sequencing data. *Methods*, 2022, 203: 478–487
- 52 McCormick C A, Meseonznik M, Qiu Y, et al. mRNA psi profiling using nanopore DRS reveals cell type-specific pseudouridylation. *bioRxiv*, 2024, 593203
- 53 Tavakoli S, Nabizadeh M, Makhmreh A, et al. Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. *Nat Commun*, 2023, 14: 334
- 54 Bansal M, Gupta A, Ding J, et al. Integrative analysis of nanopore direct RNA sequencing data reveals a global impact of pseudouridylation on m⁶A and m⁵C modifications. *npj Precis Onc*, 2026, 10: 52
- 55 Nguyen T A, Heng J W J, Kaewsapsak P, et al. Direct identification of A-to-I editing sites with nanopore native RNA sequencing. *Nat Methods*, 2022, 19: 833–844
- 56 Chen L, Ou L, Jing X, et al. DeepEdit: single-molecule detection and phasing of A-to-I RNA editing events using nanopore direct RNA

- sequencing. *Genome Biol*, 2023, 24: 75
- 57 Zhao Z, Yan W, Weng X. RNA modifications identification based on chemical reactions. *Bioorg Med Chem*, 2024, 111: 117861
- 58 Verwilt J, Mestdagh P, Vandesompele J. Artifacts and biases of the reverse transcription reaction in RNA sequencing. *RNA*, 2023, 29: 889–897
- 59 Workman R E, Tang A D, Tang P S, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*, 2019, 16: 1297–1305
- 60 Li R, Ren X, Ding Q, et al. Direct full-length RNA sequencing reveals unexpected transcriptome complexity during *Caenorhabditis elegans* development. *Genome Res*, 2020, 30: 287–298
- 61 Zhang S, Li R, Zhang L, et al. New insights into *Arabidopsis* transcriptome complexity revealed by direct sequencing of native RNAs. *Nucleic Acids Res*, 2020, 48: 7700–7711
- 62 Jiang C, Li P, Cao H. Unveiling tissue-specific RNA landscapes in mouse organs during fasting and feeding using nanopore direct RNA sequencing. *Adv Sci*, 2025, 12: e2408054
- 63 Jiang C, Li P, Ma Y, et al. Comprehensive gene profiling of the metabolic landscape of humanized livers in mice. *J Hepatol*, 2024, 80: 622–633
- 64 Arechederra M, Ávila M A, Berasain C. Liquid biopsy for cancer management: a revolutionary but still limited new tool for precision medicine. *Adv Lab Med*, 2020, 1: 20200009
- 65 Lin D, Shen L, Luo M, et al. Circulating tumor cells: biology and clinical significance. *Sig Transduct Target Ther*, 2021, 6: 404
- 66 Luo H, Wei W, Ye Z, et al. Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends Mol Med*, 2021, 27: 482–500
- 67 Wang J, Huang J, Hu Y, et al. Terminal modifications independent cell-free RNA sequencing enables sensitive early cancer detection and classification. *Nat Commun*, 2024, 15: 156
- 68 Schrag D, Beer T M, McDonnell Iii C H, et al. Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. *Lancet*, 2023, 402: 1251–1260
- 69 Relier S, Amalric A, Attina A, et al. Multivariate analysis of RNA chemistry marks uncovers epitranscriptomics-based biomarker signature for adult diffuse glioma diagnostics. *Anal Chem*, 2022, 94: 11967–11972
- 70 Amalric A, Bastide A, Attina A, et al. Quantifying RNA modifications by mass spectrometry: a novel source of biomarkers in oncology. *Crit Rev Clin Lab Sci*, 2022, 59: 1–18
- 71 Roach N P, Sadowski N, Alessi A F, et al. The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome Res*, 2020, 30: 299–312
- 72 Delgado-Tejedor A, Medina R, Begik O, et al. Native RNA nanopore sequencing reveals antibiotic-induced loss of rRNA modifications in the A- and P-sites. *Nat Commun*, 2024, 15: 10054
- 73 Hewel C, Wierzeiko A, Miedema J, et al. Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. *Nucleic Acids Res*, 2025, 53: gkaf1314
- 74 Milenkovic I, Cruciani S, Llovera L, et al. Epitranscriptomic rRNA fingerprinting reveals tissue-of-origin and tumor-specific signatures. *Mol Cell*, 2025, 85: 177–190.e7
- 75 Destefanis E, Avşar G, Groza P, et al. A mark of disease: how mRNA modifications shape genetic and acquired pathologies. *RNA*, 2021, 27: 367–389
- 76 Chen M, Wei L, Law C, et al. RNA *N*⁶-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. *Hepatology*, 2018, 67: 2254–2270
- 77 Ofengand J, Bakin A. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol*, 1997, 266: 246–268
- 78 McCloskey J A, Rozenski J. The small subunit rRNA modification database. *Nucleic Acids Res*, 2005, 33: D135–D138
- 79 Natchiar S K, Myasnikov A G, Kratzat H, et al. Visualization of chemical modifications in the human 80S ribosome structure. *Nature*, 2017, 551: 472–477
- 80 Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol*, 2012, 13: 355–369
- 81 Goodarzi H, Nguyen H C B, Zhang S, et al. Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell*, 2016, 165: 1416–1427
- 82 Rapino F, Delaunay S, Rambow F, et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature*, 2018, 558: 605–609

- 83 Pavon-Eternod M, Gomes S, Geslain R, et al. tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res*, 2009, 37: 7268–7280
- 84 Añazco-Guenkova A M, Miguel-López B, Monteagudo-García Ó, et al. The impact of tRNA modifications on translation in cancer: identifying novel therapeutic avenues. *NAR Cancer*, 2024, 6: zcae012
- 85 Zheng G, Qin Y, Clark W C, et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods*, 2015, 12: 835–837
- 86 Lucas M C, Pryszcz L P, Medina R, et al. Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat Biotechnol*, 2024, 42: 72–86
- 87 White L K, Dobson K, Del Pozo S, et al. Comparative analysis of 43 distinct RNA modifications by nanopore tRNA sequencing. *bioRxiv*, 2024, 604651
- 88 Spornraft M, Kirchner B, Haase B, et al. Optimization of extraction of circulating RNAs from plasma—enabling small RNA sequencing. *PLoS One*, 2014, 9: e107259
- 89 Xiang J F, Yang Q, Liu C X, et al. N^6 -methyladenosines modulate A-to-I RNA editing. *Mol Cell*, 2018, 69: 126–135.e6
- 90 Zhong Z D, Xie Y Y, Chen H X, et al. Systematic comparison of tools used for m^6A mapping from nanopore direct RNA sequencing. *Nat Commun*, 2023, 14: 1906
- 91 Maestri S, Furlan M, Mulrone L, et al. Benchmarking of computational methods for m^6A profiling with nanopore direct RNA sequencing. *Brief Bioinf*, 2024, 25: bbae001
- 92 Luo T, Xu M, Wang M, et al. Systematic evaluation of computational tools for multitype RNA modification detection using nanopore direct RNA sequencing. *Nat Methods*, 2026, 23: 438–450
- 93 Cozzuto L, Liu H, Pryszcz L P, et al. MasterOfPores: a workflow for the analysis of Oxford Nanopore direct RNA sequencing datasets. *Front Genet*, 2020, 11: 211

Nanopore sequencing for RNA modification detection: current algorithm development and future applications

LUO TingTing, XU MoPing, WANG Miao, CHEN FaYing & SHI JieJun*

Shanghai Tongji Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200065, China

* Corresponding author; E-mail: shij@tongji.edu.cn

More than 170 chemical modifications have been identified in RNA molecules, playing crucial roles in regulating RNA metabolism and function. Unlike traditional indirect detection methods, Oxford Nanopore direct RNA sequencing enables direct RNA reading. This technology facilitates the simultaneous detection of multiple RNA modifications in a single experiment by analyzing modification-induced disturbances in the current signal and base-calling errors through computational modeling. This review systematically summarizes the current landscape of computational methods developed for RNA modification detection using this platform. It highlights two primary methodological categories—supervised machine learning and differential comparison—for identifying key modifications such as N^6 -methyladenosine, pseudouridine, and 5-methylcytosine, while discussing their underlying principles, characteristics, and suitable applications. Furthermore, the review outlines the technology's potential in mapping transcriptome-wide modifications, elucidating their biological functions, and discovering disease biomarkers. Finally, we discuss existing challenges and future directions, particularly concerning the simultaneous detection of multiple modifications, improving model generalizability, and establishing standardized analytical frameworks.

RNA modifications, nanopore sequencing, bioinformatics algorithms, epitranscriptomics

doi: [10.1360/SSV-2025-0377](https://doi.org/10.1360/SSV-2025-0377)