

## 论文

# 基于亚细胞转录组解析拟南芥转录本剪接与染色质释放的耦合关系

叶雨彤, 彭慧琳, 林学富, 江颖荷, 董志诚, 刘敏\*

广州大学广东省植物适应性与分子设计重点实验室, 广州市作物基因编辑应用重点实验室, 分子遗传与进化创新研究中心, 广州大学生命科学学院, 广州 510006

\* 联系人, Email: minl@gzhu.edu.cn

国家自然科学基金(批准号: 32270606)资助

收稿日期: 2026-02-02; 接受日期: 2026-05-14

**摘要** 内含子剪接是真核生物基因表达调控的核心环节, 但植物中剪接加工与转录本染色质释放的动态耦合机制尚不清晰. 本研究以拟南芥(*Arabidopsis thaliana*)为材料, 通过深度测序解析了四种不同加工阶段的亚细胞 RNA 组分: 正在转录的染色质结合 RNA (Chromatin-bound non-polyadenylated RNA, CB)、转录终止后滞留于染色质的带尾 RNA (Chromatin-bound polyadenylated RNA, CB\_pA)、核内释放态 RNA (Nuclear polyadenylated RNA, Nuc\_pA) 及全细胞稳态 RNA (Total polyadenylated RNA, Tot\_pA). 结果显示, 拟南芥剪接过程随 RNA 成熟度提高逐渐完成, 且共转录剪接程度显著受内含子长度及多聚腺苷酸化位点距离的影响. 通过建立回归模型, 本研究识别出释放动力学存在显著差异的“快释放”与“慢释放”基因, 并发现两类基因在功能特征上具有明显差异. 进一步分析表明, 转录本的加工成熟度与其细胞内稳定性呈正相关. 综上所述, 本研究揭示了拟南芥中存在剪接驱动的转录后质量监控机制, 该机制协同调控转录本的染色质释放与稳态水平, 为解析植物基因表达的时空动态调控提供了新视角.

**关键词** 拟南芥 染色质结合 RNA 剪接 染色质释放

真核生物基因表达是一个受到多层次精密调控的复杂过程, 涵盖了从染色质水平的转录起始、延伸、终止, 以及转录后的 RNA 加工、核输出及胞质降解等多个环节<sup>[1-4]</sup>. 其中, 前体信使 RNA (precursor messenger RNA, pre-mRNA) 的内含子剪接(Splicing)是核心调控步骤之一<sup>[5,6]</sup>. 剪接不仅负责切除内含子以产生功能性转录本, 更通过可变剪接机制赋予了生物体在生长发育与应激响应中的高度灵活性<sup>[7,8]</sup>. 例如, 植物基因组中超过 70% 的多外显子基因存在可变剪接, 且该过程受环境信号高度诱导<sup>[9,10]</sup>.

传统视角常将剪接视为转录后的独立加工过程, 但越来越多的研究证实, 超过 70% 的剪接事件在 RNA 聚合酶 II (Pol II) 延伸过程中即已完成, 这种与转录过程偶联的现象被称为共转录剪接(Co-transcriptional splicing)<sup>[11-13]</sup>, 其效率受到 Pol II 延伸速率、染色质状态及特定组蛋白修饰的精细招募调控<sup>[14-16]</sup>. 然而, 转录的完成并不意味着剪接的终结. 近年来的研究显示, 相当比例的内含子在多聚腺苷酸化(Polyadenylation)发生后仍滞留在染色质上继续加工. 例如, 在哺乳动物中, 高达 27%~40% 的内含子在转录终止后才被去除<sup>[17,18]</sup>. 在模式植物

拟南芥中, 基于第三代高通量测序的研究也发现约 30%的染色质结合且含 poly(A)尾的转录本仍处于部分剪接状态<sup>[19]</sup>. 这一证据表明, 转录后剪接(Post-transcriptional splicing)同样是基因表达调控的关键层级, 且可能成为转录本从染色质释放进入后续环节的限速步骤.

植物由于无法移动, 必须依赖精准、快速的分子机制快速响应环境波动, 其中剪接的时空动态调节至关重要. 虽然已有研究对拟南芥剪接因子的功能及稳态转录组图谱进行了广泛探索, 揭示了可变剪接是调控植物基因表达的重要方式<sup>[9,10,20-23]</sup>. 但多数工作基于稳态 mRNA, 难以揭示剪接作为动态过程对 mRNA 产量与命运的即时影响. 目前, 在植物中, 关于内含子剪接成熟度如何调控转录本从染色质释放, 以及此类加工差异对 mRNA 最终命运的影响, 在全基因组层面仍缺乏系统性解析. 由于常规转录组测序无法区分处于不同加工瞬态的 RNA 分子, 分离并分析不同亚细胞组分的 RNA 已成为揭示这一动态过程的关键技术途径<sup>[24-27]</sup>.

本研究以拟南芥为材料, 通过对四种关键 RNA 组分(图 1A): 染色质结合未加尾 RNA (CB, 代表转录延伸阶段)、染色质结合且带尾 RNA (CB\_pA, 代表转录终止后的滞留于染色质阶段)、细胞核带尾 RNA (Nuc\_pA, 代表核内释放阶段)及总 poly(A) RNA (Tot\_pA, 代表细胞稳态水平), 结合深度测序与比较分析, 系统构建了拟南芥转录本在剪接加工与染色质释放动态过程中的全景图谱, 并探究了剪接成熟度与转录本释放及稳定性之间的关联性.

## 1 材料与方法

### 1.1 植物材料及培养条件

本研究选用哥伦比亚型拟南芥(*Arabidopsis thaliana*, ecotype Columbia-0)作为实验材料. 种子经消毒处理后, 播种于 1/2 MS 固体培养基上. 植株置于恒温培养箱中生长, 环境参数设置为: 温度(22±1)°C, 光强度 150  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ , 光周期 16 h 光照/8 h 黑暗(长日照条件). 待生长至 10 天后, 采集植株地上部分, 立即置于液氮中速冻, 随后转入-80°C超低温冰箱保存备用. 共设置 3 个生物学重复.

### 1.2 亚细胞组分分离与 RNA 提取

#### 1.2.1 细胞核分离

细胞核提取采用此前已发表的方法<sup>[25,28]</sup>. 每个生物学重复取约 1 g 组织粉末(体积约 2.5 mL), 加入 10 mL 预冷的细胞核分离缓冲液(Nuclei Isolation Buffer, NIB; 含 2 M Hexylene Glycol, 20 mM HEPES-KOH pH 7.2, 10 mM  $\text{MgCl}_2$ , 5 mM KCl, 1% [v/v] Triton X-100, 5 mM  $\beta$ -mercaptoethanol, 2  $\mu\text{g}/\text{mL}$  Pepstatin A, 2  $\mu\text{g}/\text{mL}$  Aprotinin, 1 mM PMSF). 充分混匀后, 于冰上孵育 5~10 min.

将提取液通过铺设在 50 mL 离心管口的两层尼龙筛网(上层 100  $\mu\text{m}$ , 下层 40  $\mu\text{m}$ )进行过滤, 并使用 NIB 将滤液补充至 15 mL.

在另一 50 mL 离心管中建立不连续 Percoll 密度梯度: 先加入 6 mL 80% Percoll 溶液, 随后在其上方极其轻缓地铺设 6 mL 30% Percoll 溶液(梯度缓冲液成分均为 0.5 M Hexylene

Glycol, 5 mM HEPES-KOH pH 7.2, 10 mM MgCl<sub>2</sub>, 5 mM KCl, 1% [v/v] Triton X-100, 5 mM β-mercaptoethanol), 确保 80% 与 30% 溶液的交界面清晰无扰动。

将前述 15 mL 滤液小心加至 30% Percoll 层之上。使用水平转子离心机在 4 °C, 2,000 × g 条件下离心 11 min, 设置最低升速与降速水平(以防扰动梯度)。离心后, 弃去上层 NIB, 使用移液管小心收集处于 30%~80% 界面处的细胞核富集层。

向收集的细胞核中加入预冷的梯度缓冲液至 15 mL, 在 4 °C, 2,000 × g 条件下离心 6 min。弃上清, 使用 500 μL 预冷的梯度缓冲液重悬。最后将细胞核悬液转移至新的 1.5 mL 离心管中, 于 4 °C, 8,000 × g 条件下离心 3 min 收集细胞核沉淀。

### 1.2.2 染色质提取

向纯化的细胞核沉淀中加入 150 μL 预冷的细胞核重悬缓冲液(Nuclei Resuspension Buffer, NRB; 含 50% Glycerol, 0.5 mM EDTA, 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT, 0.4 U/μL RNase inhibitor, 2 μg/mL Pepstatin A, 2 μg/mL Aprotinin, 1 mM PMSF, 8 mM β-mercaptoethanol), 轻轻吹打重悬。随后加入 300 μL 洗涤缓冲液(Washing Buffer, WB; 含 25 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 M Urea, 0.5 mM EDTA, 1% Tween 20, 0.4 U/μL RNase inhibitor, 1 mM DTT, 2 μg/mL Pepstatin A, 2 μg/mL Aprotinin, 2 mM PMSF, 8 mM β-mercaptoethanol), 充分混匀。在 4 °C, 8,000 × g 条件下离心 1 min, 弃去上清液。重复上述步骤一次: 即加入 150 μL NRB 重悬沉淀后, 再加入 150 μL WB 混匀, 于 4 °C, 8,000 × g 离心 1 min。彻底弃去上清液后, 于 4 °C, 12,000 × g 短促离心 2 min, 使用微量移液器吸除残留的上清液, 所得沉淀即为纯化的染色质组分。

### 1.2.3 RNA 提取

本研究分别提取了总 RNA、细胞核 RNA 及染色质结合 RNA。各组分的起始材料量分别为: 总 RNA 提取采用约 100 μL 组织粉末; 细胞核 RNA 提取采用约 50 μL 纯化后的细胞核沉淀; 染色质 RNA 提取则采用前述步骤获得的全部染色质组分。

具体提取步骤如下: 向上述材料中各加入 0.7 mL TRIzol 试剂, 剧烈振荡混匀后, 于室温下静置 5~10 min 以充分裂解。随后加入 200 μL 氯仿, 振荡 30 s, 并室温静置 2~3 min。于 4°C, 12000 × g 离心 15 min。小心吸取上层水相至新的 RNase-free 离心管中。为提高 RNA 纯度, 重复氯仿抽提一次: 加入 400 μL 氯仿, 并重复上述振荡与离心步骤, 再次收集上层水相。向收集的水相中加入 3 倍体积预冷的无水乙醇, 1/10 倍体积的 5 M 醋酸铵溶液及 1 μL GlycoBlue 共沉淀剂, 颠倒混匀后, 置于 -80°C 下沉淀 2 小时以上。沉淀结束后, 于 4°C, 12000 × g 离心 15 min, 弃上清。使用 1 mL 预冷的 75% 乙醇洗涤沉淀两次。室温下开盖晾干残留乙醇, 加入 16 μL RNase-free water 充分溶解 RNA 沉淀。利用 Qubit 荧光定量仪测定 RNA 浓度后, 用于后续文库构建。

## 1.3 文库构建及测序

利用 TURBO DNase (Life Technologies) 彻底降解各组分中的基因组 DNA。随后对 CB

RNA 组分进行如下处理: 首先使用 Pan-Plant riboPOOL 试剂盒(siTOOLS Biotech)去除其中的核糖体 RNA (rRNA); 接着利用 Oligo d(T)<sub>25</sub> 磁珠(NEB, S1419)去除其中 poly(A) RNA, 从而获得纯净的无尾 CB RNA. 与此同时, 将 Oligo d(T)<sub>25</sub> 磁珠从 CB RNA 中富集的含 poly(A)尾 RNA 洗涤下来, 即获得 CB\_pA RNA.

对于细胞核和总 RNA, 采用 oligo(dT)磁珠从中富集的含 poly(A)尾 RNA, 即获得 Nuc\_pA 和 Tot\_pA RNA.

最后, 使用 NEBNext Ultra II 方向性 RNA 文库构建试剂盒(NEB, E7765)分别将四种转化为 cDNA 文库, 并依托 Illumina NovaSeq 测序平台进行高通量测序.

## 1.4 生物信息学分析

### 1.4.1 转录组数据处理

首先, 利用 Cutadapt 软件<sup>[29]</sup>去除原始数据中的接头序列及低质量读段. 随后, 通过 STAR 软件<sup>[30]</sup>将质控后的 clean reads 比对到拟南芥基因组(TAIR10), 仅保留唯一比对序列. 最后, 采用 featureCounts<sup>[31]</sup>统计基因上的 reads 数目, 并使用 R 软件<sup>[32]</sup>计算基因的表达量 TPM (Transcript Per Million), 及进行相关性和主成分分析(Principal Component Analysis, PCA).

染色质结合 RNA 中, 除了正在转录的 RNA 之外, 还存在结合在染色质上起调控作用的 RNA, 为了排除这部分 RNA 的干扰, 后续分析仅针对蛋白编码基因进行.

### 1.4.2 内含子与外显子比值计算

本研究通过计算内含子与外显子的测序 reads 密度比值(Intron/Exon Ratio, IER)来评估转录本的成熟度. 首先基于 Araport11 注释文件, 提取每个基因所有转录本共有的内含子区域作为该基因的内含子参考, 其余部分视为外显子. 利用 featureCounts 分别统计内含子与外显子区域的 reads 数, 并根据区域长度换算为 reads 密度. IER 计算方式为内含子 reads 密度除以外显子 reads 密度.

### 1.4.3 PIR 指数计算

为了排除可变剪接的干扰, 本研究针对组成型剪接(Constitutive Splicing)的内含子, 即没有可变剪接的内含子, 计算其内含子保留指数(Percent Intron Retained, PIR)<sup>[33]</sup>. 分析流程参考已发表文章中的自写脚本<sup>[25]</sup>. 具体为, 针对每个内含子统计三类比对序列: (1) 跨越 5'剪接位点的外显子-内含子衔接序列(exon-intron reads at 5'SS, EI5); (2) 跨越 3'剪接位点的外显子-内含子衔接序列(exon-intron reads at 3'SS, EI3); (3) 剪接后的外显子-外显子衔接序列(exon-exon junction reads, EE)(图 3A). PIR 计算公式如下:

$$\text{PIR} = (\text{EI5} + \text{EI3}) / ((\text{EI5} + \text{EI3}) + 2 \times \text{EE})$$

### 1.4.4 线性回归分析和统计检验

本研究为量化不同基因的释放动力学特征, 以及转录本加工成熟度与稳定性的关系, 基

于各基因在 CB、CB\_pA 和 Tot\_pA 中的表达量( $\log_2$  TPM)进行分析, 以 CB 水平的  $\log_2$  TPM 为自变量, CB\_pA 和 Tot\_pA 水平的  $\log_2$  TPM 分别为因变量, 利用 R 软件中的 `lm()` 函数进行线性回归. 为识别具有显著释放差异及稳定性差异基因, 参考差异表达分析中常用的倍数变化阈值, 将观测值与预测值偏离程度超过 2.5 倍或低于 0.4 倍的基因定义为极端类型. 在 CB 与 CB\_pA 的回归中, 以回归线为基准, 将 CB\_pA 实际 TPM 值高于预测值 2.5 倍的基因定义为慢释放基因(Slow-release genes), 低于预测值 0.4 倍的基因定义为快释放基因(Fast-release genes), 其余为中速释放基因(Medium-release genes). 在 CB 与 Tot\_pA 的回归中, 将 Tot\_pA 实际 TPM 值高于预测值 2.5 倍的基因定义为稳定(Stable)的基因, 低于预测值 0.4 倍的基因定义为不稳定(Unstable)的基因.

文中所涉及的统计检验, 如 Wilcoxon 秩和检验(Wilcoxon rank-sum test)和 Spearman 秩相关系数(Spearman's rank correlation coefficient)计算均在 R 软件中完成.

#### 1.4.5 GO 富集分析

为探究不同类型基因的生物学功能, 我们采用 `clusterProfiler` 软件包<sup>[34]</sup>对其进行了 GO (Gene Ontology)富集分析, 聚焦于生物学过程(Biological Process, BP), 显著性阈值设为校正后  $P$  值  $< 0.05$  (FDR 校正).

## 2 结果与分析

### 2.1 不同 RNA 组分的相关性分析与主成分分析

为保证数据分析的可靠性, 首先对各 RNA 组分样本的基因表达水平进行了生物学重复间的一致性评估. 结果显示, 所有样本的生物学重复之间均表现出高度的一致性, Pearson 相关系数均在 0.97 至 0.99 之间(图 1B). PCA 结果进一步揭示, CB 与 CB\_pA 样本具有明显差异, 且与 Nuc\_pA 及 Tot\_pA 样本显著分离, 反映了不同亚细胞定位的转录组在组成上存在本质差异. 相比之下, Nuc\_pA 与 Tot\_pA 在 PCA 空间中聚集紧密, 提示二者包含的转录本组成高度相似. 鉴于生物学重复表现出优异的重现性, 后续分析将合并重复样本以提高定量精度.

本研究所产生的转录组学数据已上传至国家基因组科学数据中心 (<https://ngdc.cncb.ac.cn/>; BioProject: PRJCA060455).

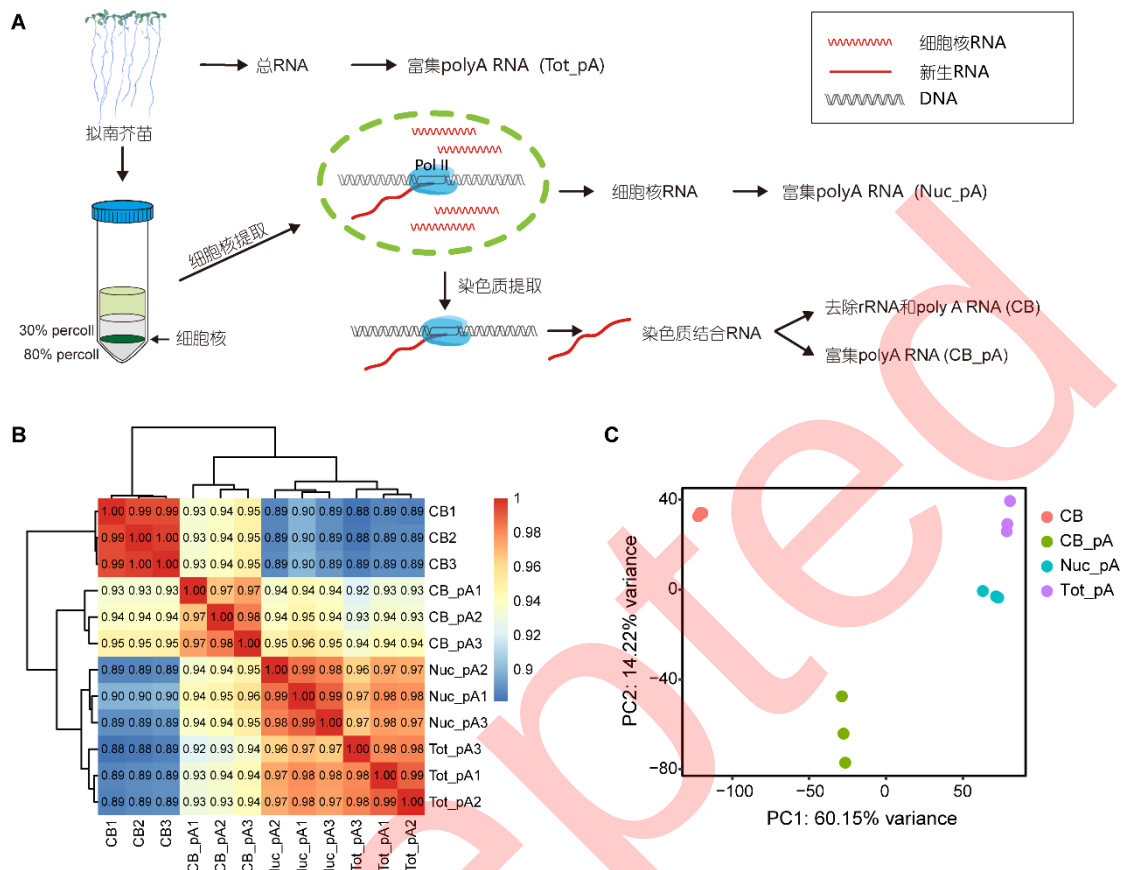


图1 亚细胞组分 RNA 分离流程及生物学重复的一致性评估. A: 不同组分 RNA 的分离与纯化流程示意图; B: 各样本生物学重复间的相关性热图及层次聚类. 颜色及深浅代表 Pearson 相关系数的高低, 深红色表示相关系数高, 深蓝色表示相关系数相对低, 数值标注于方格内; C: 不同转录组样本的主成分分析(PCA)图. 图中不同颜色的点代表不同组分的 RNA 样本.

Figure 1 Workflow for RNA isolation from subcellular fractions and assessment of biological replicates. A: Schematic diagram of the process for subcellular fractionation RNA isolation and purification; B: Heatmap of correlations and hierarchical clustering among biological replicates of each sample. Colors and shades indicate the level of Pearson correlation coefficients, with dark red representing high correlation and dark blue representing relatively low correlation. Numerical values are annotated within the squares; C: Principal component analysis (PCA) plot of different transcriptomic samples. Points in different colors represent RNA samples from different components.

## 2.2 内含子剪接成熟度随 RNA 加工进程而提高

根据组分定义, CB、CB\_pA、Nuc\_pA 和 Tot\_pA 依次代表了从新生转录、染色质滞留、核内释放到细胞稳态的四个关键进程. 理论上, 随 RNA 加工进程的推进, 未被剪除的内含子比例应逐步下降.

为此, 我们计算了各组分中每个基因的内含子与外显子覆盖度比值(Intron/Exon ratio, IER). 整体分析显示, 该比值在 4 个组分间呈显著递减趋势(图 2A). CB 组分 IER 值最高, 反映了其富含未剪接转录本的特征; CB\_pA 次之, 表明部分内含子转录终止后剪接仍在持续; 而 Tot\_pA 降至最低, 说明其主要由成熟 mRNA 组成. 此外, 通过 IGV 基因组浏览器对单个基因的观察发现(图 2B-D), 随着组分成熟度提高, 内含子区域的测序信号逐渐消失, 外显子连接趋于连续. 这一单基因水平的证据与全局分析结果一致, 系统地证实了剪接随 RNA 加

工逐步完成的动态过程。

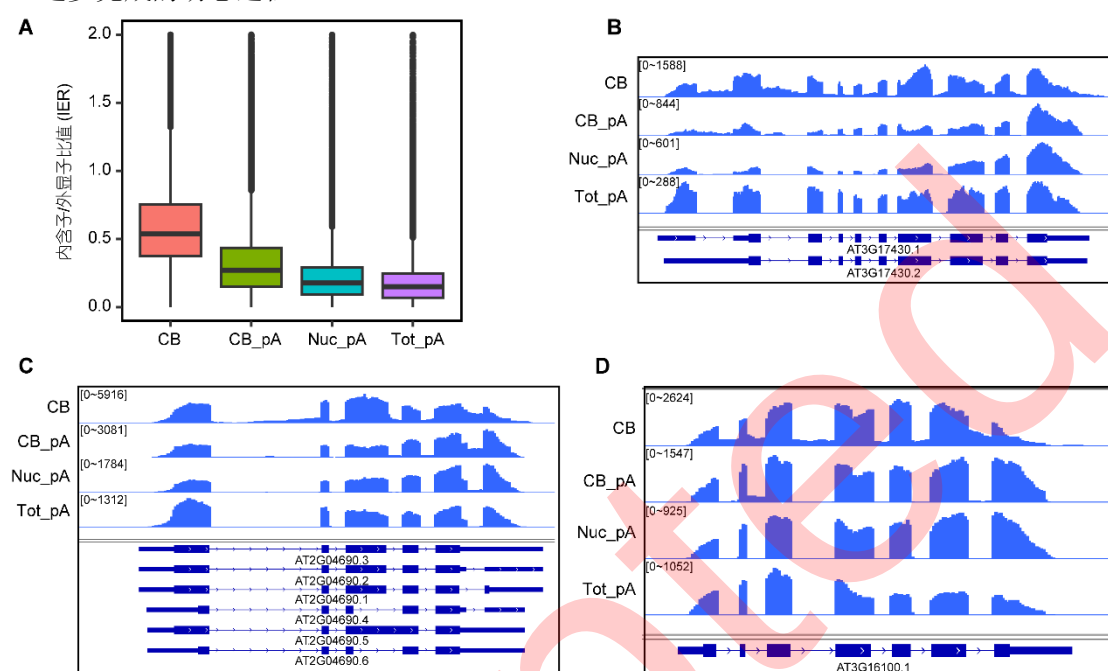


图2 不同组分 RNA 内含子与外显子比例比较. A: 不同 RNA 组分中内含子与外显子测序 reads 密度的比值(Intron/Exon ratio, IER)分布; B-D: 3 个基因剪接动态的 IGV 可视化展示实例.

Figure 2 Comparison of intron-to-exon ratios in RNA from different subcellular fractions. A: Distribution of the intron-to-exon read density ratio (IER) in RNA from different subcellular fractions; B-D: Visualization examples showing splicing dynamics of three genes in IGV.

### 2.3 不同阶段内含子剪接特征及其影响因素

为进一步探究内含子剪接效率的影响因素,本研究聚焦于组成型剪接的内含子,利用内含子保留指数(PIR)对其剪接成熟度进行定量评估. PIR 值越低,代表该内含子的剪接程度越完全(图 3A).

结果显示, PIR 值在 CB、CB\_pA、Nuc\_pA 至 Tot\_pA 的组分中呈现显著的阶梯式下降趋势(图 3B). 这与 IER 的整体变化规律一致,从剪接位点水平进一步证实了剪接随 RNA 加工进程逐步完善的动态特征.

进一步分析发现,内含子长度及其与多聚腺苷酸化位点(polyadenylation site, PAS)的距离是影响共转录剪接效率的关键因素. 在 CB 组分中, PIR 与内含子长度呈正相关(Spearman 相关系数  $\rho = 0.180$ ,  $P < 2.2 \times 10^{-16}$ ),即较长内含子的剪接效率较低(图 3C);同时, PIR 随内含子与 PAS 距离的缩短而升高( $\rho = -0.212$ ,  $P < 2.2 \times 10^{-16}$ ),符合“先转录,先剪接”的顺次加工规律(图 3D),与此前的研究一致<sup>[26]</sup>. 而在转录终止后仍滞留于染色质的 CB\_pA 组分中,虽然 PIR 仍随内含子长度增加而升高( $\rho = 0.142$ ,  $P < 2.2 \times 10^{-16}$ ),但其与 PAS 距离的相关性则呈现与 CB 组分相反的趋势,为正相关( $\rho = 0.127$ ,  $P < 2.2 \times 10^{-16}$ ),即距离 PAS 越远的内含子 PIR 值越高. 这表明,在加尾完成后持续的染色质滞留,可能为末端内含子的剪接提供了额外的补偿时间,而最早转录出来的一些内含子可能更依赖于脱离染色质的转录后剪接加工.

至核内释放(Nuc\_pA)及细胞稳态(Tot\_pA)阶段,绝大多数内含子已完成剪接,整体 PIR

降至极低水平, 仅较长内含子仍表现出略高的 PIR 值. 尽管该阶段 PIR 与内含子长度及 PAS 距离的相关性仍具有统计学显著性( $P < 0.05$ ), 但相关系数绝对值已低于 0.05, 反映了加工后期剪接事件的高度完成性.

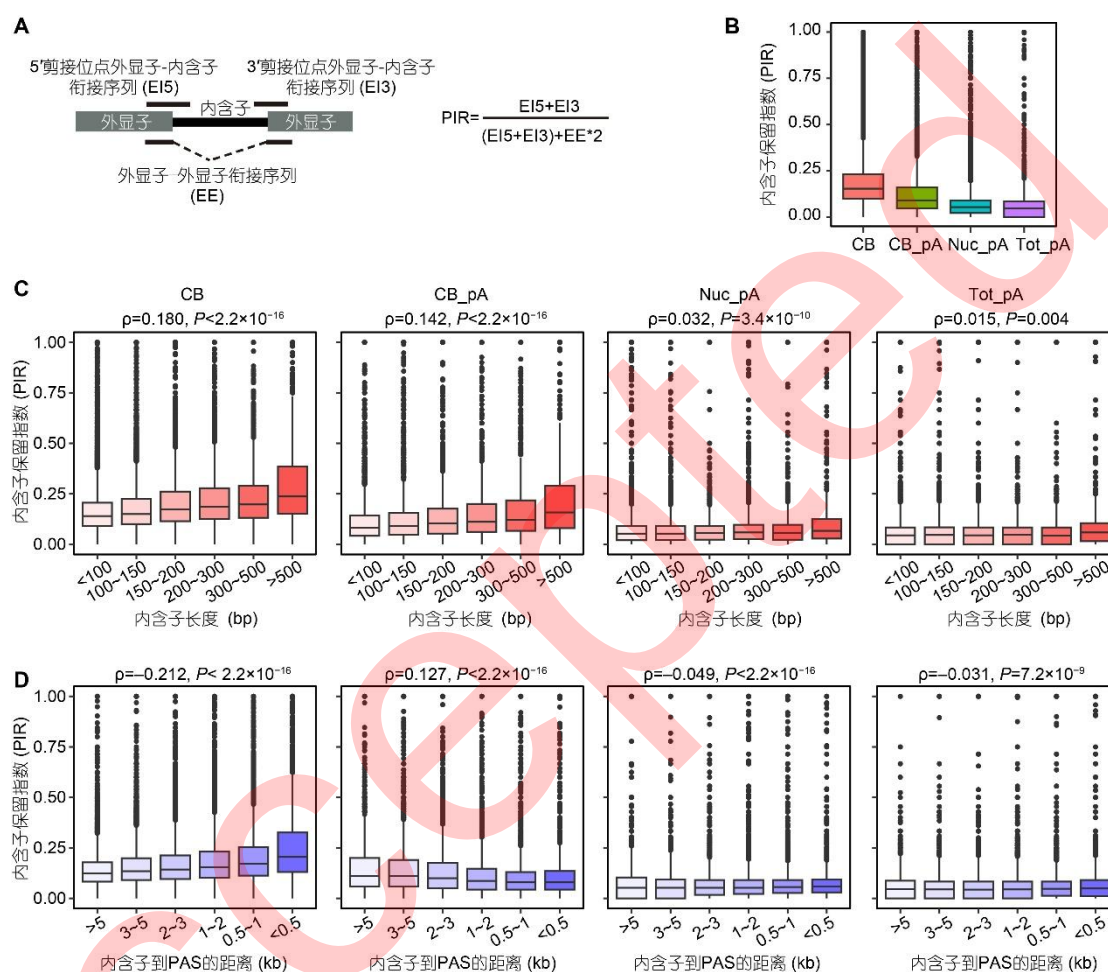


图3 内含子剪接效率的量化及其影响因素分析. A: 内含子保留指 PIR 计算方法示意图; B: 四种 RNA 组分(CB, CB\_pA, Nuc\_pA, Tot\_pA)组成型内含子的 PIR 分布箱线图. PIR 值越低表示剪接完成度越高; C: 不同长度内含子的 PIR 指数比较; D: 内含子距离 PAS 不同距离的内含子 PIR 指数比较. C 和 D 的图上上方显示了 Spearman 秩相关检验的相关系数  $\rho$  和  $P$  值.

Figure 3 Quantification of intron splicing efficiency and analysis of its influencing factors. A: Schematic diagram of the calculation method for the intron retention index (PIR); B: Box plots showing the PIR distribution of constitutive introns in four RNA components (CB, CB\_pA, Nuc\_pA, Tot\_pA). Lower PIR values indicate higher splicing completion; C: Comparison of PIR indices for introns of different lengths; D: Comparison of PIR indices for introns at varying distances from the PAS. The Spearman's rank correlation coefficient  $\rho$  and  $P$  value are shown above panels C and D.

## 2.4 内含子加工效率与转录本的染色质释放动力学相关

CB 组分反映了基因的转录活性, 而 CB\_pA 组分则代表已完成多聚腺苷酸化但尚未从染色质释放的转录本. 当转录与释放达到稳态平衡时, 基因在两组分中的相对丰度(TPM)理论上应呈现高度相关. 但实际中, 观测到的相关性受多种因素影响, 如基因长度、测序深度、

细胞组分分离效率及建库引入的偏差等。虽然 TPM 在计算时已经对长度和测序深度进行了标准化处理, 但上述其他因素仍可能对定量结果产生一定干扰。尽管如此, 观测值相对于回归分析预测值的偏离程度仍可为量化不同基因的释放动力学特征提供有效依据。对基因在 CB 和 CB\_pA 水平的  $\log_2$  TPM 值进行回归, 结果显示, 两者呈显著正相关( $R^2 = 0.752$ ,  $P < 0.001$ )(图 4A)。以该回归线为基准, 依据实际值偏离预测值的程度(阈值设置详见材料与方

法), 共识别出慢释放基因 1,908 个、快释放基因 2,896 个, 其余定义为中速释放基因(图 4A)。我们进一步分析了释放速率与内含子加工状态的关系(图 4B)。在 CB 水平(共转录阶段), 快释放基因的内含子/外显子比值最低, 而慢释放基因最高, 表明共转录剪接效率与释放速度呈正相关。有趣的是, 在 CB\_pA 水平(转录后滞留阶段), 趋势发生逆转: 慢释放基因的内含子占比显著降低, 甚至低于快释放基因。这表明慢释放基因虽在转录阶段剪接较慢, 但通过在染色质上较长时间的滞留, 最终在释放前完成了更高比例的内含子切除, 即存在明显的转录后剪接补偿。

GO 功能富集分析揭示了快、慢释放基因在生物学功能上的明显分化。慢释放基因显著富集于氧化胁迫响应、蛋白质折叠及翻译相关过程(图 4C); 而快释放基因则显著富集于 DNA 损伤修复、细胞周期、RNA 修饰及蛋白质磷酸化等功能模块(图 4D)。这表明, 转录本的染色质释放速率可能是一种调控机制, 通过控制不同功能类别 mRNA 的产出时效性, 以适应细胞不同的生理需求。



共 3,503 个; 将观测值低于预测值 0.4 倍的基因定义为具有不稳定转录本的基因, 共 4,784 个; 其余定义为常规基因(图 5A).

通过比较这三类基因在各加工阶段的内含子/外显子比值发现(图 5B), 稳定转录本基因在从新生转录本到成熟 mRNA 的所有阶段中, 其 IER 值始终维持在最低水平. 这表明转录本的加工成熟度与其在细胞内的稳定性显著正相关: 剪接更完全、加工更彻底的转录本倾向于更稳定地积累; 反之, 加工不完全的转录本则可能更容易被细胞内的质量监控机制识别并降解.

GO 功能富集分析进一步揭示了不同稳定性转录本的功能倾向(图 5C, D). 稳定转录本显著富集于响应氧化胁迫、次生代谢过程、质体组织及细胞壁组织等生物学途径; 而不稳定转录本则主要富集于 DNA 损伤响应、蛋白质磷酸化及细胞周期等过程. 这表明不同功能类别的基因在进化过程中可能选择了差异化的加工效率与稳定性策略, 以满足植物对环境响应和发育调控的需要.

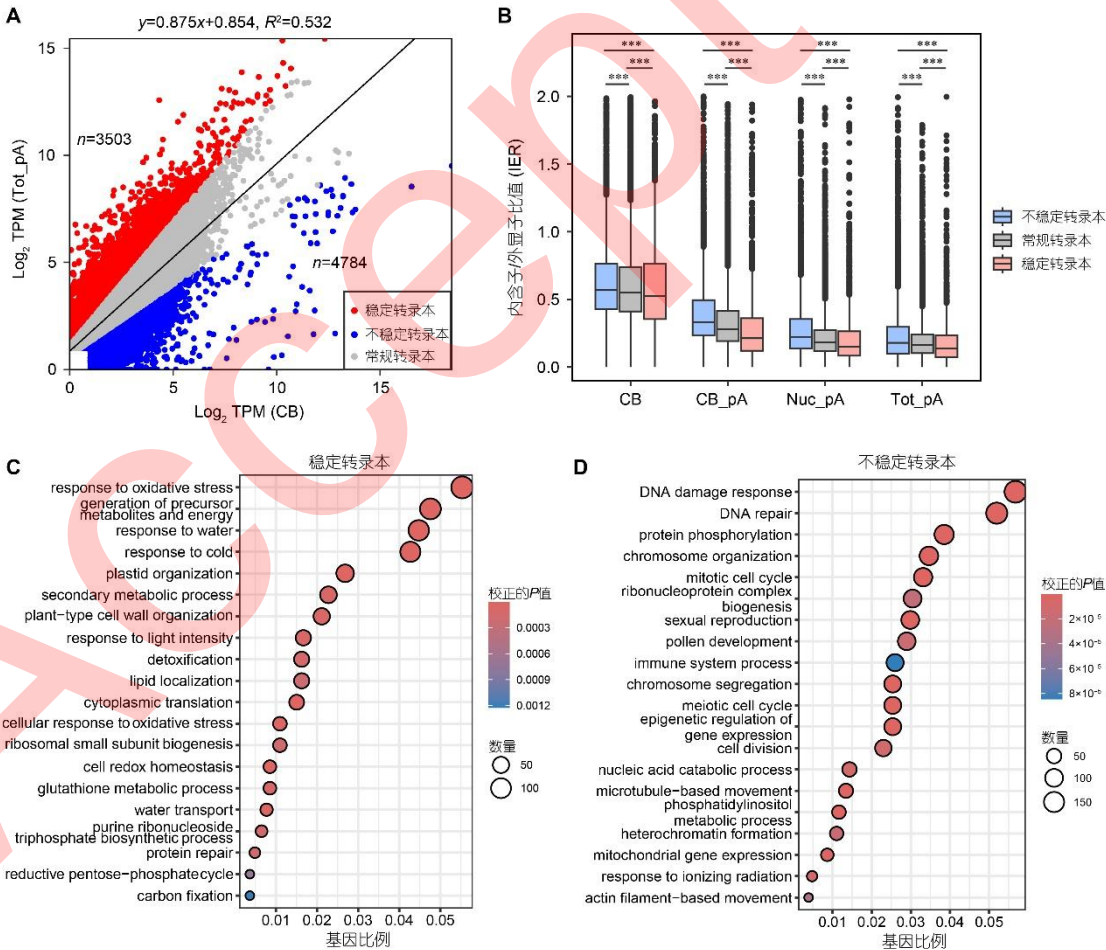


图 5 基因转录本稳定性与内含子加工程度的关系分析. A: 基于转录活性(CB)与成熟 mRNA 丰度(Tot\_pA)的转录本稳定性评估. 散点图展示了各基因在 CB 与 Tot\_pA 水平的  $\log_2$  TPM 线性回归分析. 红色点代表稳定转录本, 蓝色点代表不稳定转录本, 灰色点代表常规转录本; B: 三类稳定性基因在不同 RNA 组分中的内含子/外显子比值分布; \*\*\*,  $P < 0.001$  (Wilcoxon 秩和检验); C: 具有稳定转录本基因的 GO 功能富集分析; D: 具有不稳定转录本基因的 GO 功能富集分析.

Figure 5 Analysis of the relationship between gene transcript stability and intron processing extent. A: Assessment of transcript stability based on transcriptional activity (CB) and mature mRNA abundance (Tot\_pA). The scatter plot depicts the linear regression analysis of  $\log_2$  TPM levels for each gene in CB and Tot\_pA. Red dots represent stable transcripts, blue dots represent unstable transcripts, and gray dots represent normal transcripts; B: Distribution of intron/exon ratios for the three stability categories across different RNA components; \*\*\*,  $P < 0.001$  (Wilcoxon Rank-Sum Test); C: GO functional enrichment analysis of genes with stable transcripts; D: GO functional enrichment analysis of genes with unstable transcripts.

### 3 讨论

#### 3.1 剪接是贯穿转录与转录后的连续加工过程

在动物系统中, 剪接的时空动态已经被广泛研究, 普遍认为大部分剪接事件与染色质环境紧密耦合, 甚至转录后剪接也常与染色质关联状态下进行<sup>[17,18,35,36]</sup>. 但在植物中, 剪接加工与转录本染色质释放的动态关系尚不明确. 本研究发现, 拟南芥的新生转录本(CB 组分)虽富含内含子信号, 但已发生显著剪接; 而在转录终止后仍滞留于染色质的转录本(CB\_pA组分)中, 剪接完成度得到进一步提升, 且 Nuc\_pA 中的剪接程度已接近 Tot\_pA 水平, 表明植物中内含子剪接加工过程也基本与染色质关联. 同时拟南芥中, 共转录剪接与转录后剪接是处于一个动态演进的连续体中. 这种时空上的连续性可能为植物提供了一个转录后层面的精细调控窗口, 使其能够根据细胞状态灵活调节成熟 mRNA 的产出效率与时机.

#### 3.2 染色质滞留作为低效剪接的补偿与质控窗口

本研究通过回归模型鉴定的“慢释放基因”, 为理解 RNA 加工的监控机制提供了直接线索. 分析显示, 这类基因在共转录阶段剪接效率较低, 但在随后的染色质滞留阶段(CB\_pA)经历了显著的剪接成熟(图 4B). 这表明, 染色质不仅是转录的模板, 同时也充当了一个加工平台, 为共剪接效率较低的转录本提供了必需的补偿时间. 这种延迟释放机制有效保证了前体 RNA 在进入核质并后续转运前, 能够完成必要的成熟加工.

更重要的是, 这种滞留可能构成一个动态的质量筛选环节. 能够在滞留期内完成正确剪接的转录本可以被释放; 而未能正确加工的异常转录本, 则可能通过招募核外切酶体(Nuclear exosome)进行就地降解, 从而从源头上避免了含有内含子的缺陷型转录本进入胞质翻译环节<sup>[18,37]</sup>. 因此, 基于释放动力学的调控, 使植物能够在转录水平之外, 通过对剪接速率的微调, 实现对基因表达输出的二次校准, 这在需要快速调整基因表达以应对环境变化的场景中尤为重要.

#### 3.3 剪接效率与转录本代谢命运的关联

本研究发现, 高稳定性转录本在所有加工阶段均表现出更高的剪接完成度(图 5B), 暗示剪接效率与转录本稳定性之间存在内在关联. 这可能是因为高效剪接能更早地促进外显子连接复合物(Exon junction complex, EJC)的及时装配, 或影响了核糖核蛋白

(Ribonucleoprotein, RNP)的组成,从而增强了 mRNA 的结构稳定性和抗降解能力<sup>[38-40]</sup>. GO 富集分析进一步证实,稳定转录本多涉及氧化应激等基础代谢过程,需维持较高的稳态丰度;而涉及细胞周期及信号转导的基因剪接效率相对较低,这种较低的加工效率可能与其较快的周转速率协同,便于细胞在不同发育阶段或环境信号间实现快速的转录本切换.

### 3.4 总结与展望

本研究系统揭示了在拟南芥中,内含子剪接是一个与转录本染色质释放动力学紧密耦合的动态过程.基因表达调控不仅发生在转录起始水平,转录本的加工成熟度及其从染色质的释放速率,共同决定了 mRNA 的最终稳态丰度与功能时效性.这一剪接驱动的转录后质量监控机制,为理解植物如何协调基因表达的速度与精度提供了新视角.

本研究结论主要基于群体水平的静态组分分析,且受限于二代测序的短读长特性,得出的规律尚不够精细.未来的研究可借助体内新生 RNA 标记技术与高通量长读长测序(如 Nanopore)相结合的策略,在单分子水平实时解析特定内含子的剪接顺序、动力学及其与 Pol II 转录进程的精确关系.此外,探究环境信号如何通过影响剪接因子或改变染色质状态来调节这一释放质量控制通路,将有助于深入理解植物适应复杂环境的转录后调控基础.

## 4. 结论

本研究通过系统比较拟南芥不同亚细胞组分 RNA,揭示了内含子剪接、转录本染色质释放与 mRNA 稳定性之间的内在联系与协同调控关系.结果表明,剪接是一个从共转录延伸至转录后滞留阶段的连续动态过程,大部分剪接过程与染色质耦连.染色质滞留能够为共转录剪接效率较低的转录本提供关键的加工补偿时间.同时,转录本的剪接成熟度与其在细胞内的稳定性呈正相关,表明加工效率可能影响 mRNA 的命运与功能时长.这些结果表明,剪接不仅是生成成熟 RNA 的步骤,更是调控基因表达产量和时效的重要环节,为我们理解植物如何精准控制基因表达提供了新的依据.

## 参考文献

- 1 Cramer P. Organization and regulation of gene transcription. *Nature*, 2019, 573: 45–54
- 2 Shine M, Gordon J, Schäfer L, et al. Co-transcriptional gene regulation in eukaryotes and prokaryotes. *Nat Rev Mol Cell Biol*, 2024, 25: 534–554
- 3 Shi T, Liu Y H, Li S S. Non-canonical tailing of eukaryotic RNA. *Sci Sin Vitae*, 2024, 54: 1129–1143  
[石婷, 刘毅涵, 李思思. 真核生物 RNA 的非经典加尾修饰. *中国科学: 生命科学*, 2024, 54: 1129 – 1143]
- 4 Jia Z, Liu J W, Lu F L. Technologies for measuring RNA poly(A) tails. *Sci Sin Vitae*, 2025, 55: 1134–1148  
[贾铮, 刘静雯, 陆发隆. Poly(A)尾检测技术研究进展. *中国科学: 生命科学*, 2025, 55: 1134 – 1148]
- 5 Wilkinson M E, Charenton C, Nagai K. RNA splicing by the spliceosome. *Annu Rev Biochem*, 2020, 89: 359–388
- 6 Wan R X. Advances in the molecular mechanism of mRNA splicing by the spliceosome. *Sci Sin Vitae*, 2025, 55: 1542–1556  
[万蕊雪. 剪接体介导 mRNA 剪接的分子机理研究进展. *中国科学: 生命科学*, 2025, 55: 1542–1556]
- 7 Steward R A, De Jong M A, Oostra V, et al. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat Commun*, 2022, 13: 755
- 8 Tognacca R S, Rodríguez F S, Aballay F E, et al. Alternative splicing in plants: current knowledge and future directions for assessing the biological relevance of splice variants. *J Exp Bot*, 2023, 74: 2251–2272
- 9 Martín G, Márquez Y, Mantica F, et al. Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biol*, 2021, 22: 35
- 10 Syed N H, Kalyna M, Marquez Y, et al. Alternative splicing in plants-coming of age. *Trends Plant Sci*, 2012, 17: 616–623
- 11 Zhu D L, Mao F, Tian Y C, et al. The features and regulation of co-transcriptional splicing in *Arabidopsis*. *Mol Plant*, 2020, 13: 278–294
- 12 Tilgner H, Knowles D G, Johnson R, et al. Deep sequencing of subcellular RNA fractions shows splicing to be

- predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res*, 2012, 22: 1616–1625
- 13 Bentley D L. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet*, 2014, 15: 163–175
  - 14 Moehle E A, Braberg H, Krogan N J, et al. Adventures in time and space: splicing efficiency and RNA polymerase II elongation rate. *RNA Biol*, 2014, 11: 313–319
  - 15 Matveeva E A, Al-Tinawi Q M H, Rouchka E C, et al. Coupling of PARP1-mediated chromatin structural changes to transcriptional RNA polymerase II elongation and cotranscriptional splicing. *Epigenet Chromatin*, 2019, 12: 15
  - 16 Neves L T, Douglass S, Spreafico R, et al. The histone variant H2A.Z promotes efficient cotranscriptional splicing in *S. cerevisiae*. *Genes Dev*, 2017, 31: 702–717
  - 17 Choquet K, Baxter-Koenigs A R, Dülk S L, et al. Pre-mRNA splicing order is predetermined and maintains splicing fidelity across multi-intronic transcripts. *Nat Struct Mol Biol*, 2023, 30: 1064–1076
  - 18 Choquet K, Patop I L, Churchman L S. The regulation and function of post-transcriptional RNA splicing. *Nat Rev Genet*, 2025, 26: 378–394
  - 19 Jia J B, Long Y P, Zhang H, et al. Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat Plants*, 2020, 6: 780–788
  - 20 Chang P, Hsieh H Y, Tu S L. The U1 snRNP component RBP45d regulates temperature-responsive flowering in *Arabidopsis*. *Plant Cell*, 2022, 34: 834–851
  - 21 Li Y, Guo Q H, Liu P, et al. Dual roles of the serine/arginine-rich splicing factor SR45a in promoting and interacting with nuclear cap-binding complex to modulate the salt-stress response in *Arabidopsis*. *New Phytol*, 2021, 230: 641–655
  - 22 Alhabsi A, Ling Y, Crespi M, et al. Alternative splicing dynamics in plant adaptive responses to Stress. *Annu Rev Plant Biol*, 2025, 76: 687–717
  - 23 Duque P. A role for SR proteins in plant stress responses. *Plant Signaling Behav*, 2011, 6: 49–54
  - 24 Zhu J, Liu M, Liu X, et al. RNA polymerase II activity revealed by GRO-seq and pNET-seq in *Arabidopsis*. *Nat Plants*, 2018, 4: 1112–1123
  - 25 Liu M, Zhu J, Huang H, et al. Comparative analysis of nascent RNA sequencing methods and their applications in studies of cotranscriptional splicing dynamics. *Plant Cell*, 2023, 35: 4304–4324
  - 26 Zhu J, Zhao H, Kong F, et al. Cotranscriptional and posttranscriptional features of the transcriptome in soybean shoot apex and leaf. *Front Plant Sci*, 2021, 12: 649634
  - 27 Mayer A, di Iulio J, Maleri S, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*, 2015, 161: 541–554
  - 28 Chen Y, Zhu J, Xie Y, et al. Protocol for affordable and efficient profiling of nascent RNAs in bread wheat using GRO-seq. *STAR Protoc*, 2022, 3: 101657
  - 29 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*, 2011, 17: 10–12
  - 30 Dobin A, Davis C A, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 2013, 29: 15–21
  - 31 Liao Y, Smyth G K, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2014, 30: 923–930
  - 32 R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
  - 33 Braunschweig U, Barbosa-Morais N L, Pan Q, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*, 2014, 24: 1774–1786
  - 34 Yu G, Wang L G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 2012, 16: 284–287.
  - 35 Girard C, Barbosa-Morais N L, Pan Q, et al. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun*, 2012, 3: 994
  - 36 Brody Y, Neufeld N, Bieberstein N, et al. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLOS Biol*, 2011, 9: e1000573
  - 37 Szczepińska T, Kalisiak K, Tomecki R, et al. DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res*, 2015, 25: 1622–1633
  - 38 Hir H L, Saulière J, Wang Z. The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol*, 2016, 17: 41–54
  - 39 Abshire E T, Maquat L E. Gene regulation through exon junction complex modularity. *Nat Struct Mol Biol*, 2025, 32: 2387–2397
  - 40 Wahl M C, Will C L, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*, 2009, 136: 701–718

# Subcellular transcriptomics reveals the coupling of transcript splicing and chromatin release in *Arabidopsis*

Yutong Ye, Huilin Peng, Xuefu Lin, Yinghe Jiang, Zhicheng Dong, Min Liu\*

School of Life Sciences, Innovative Center of Molecular Genetics and Evolution, Guangzhou Key Laboratory of Crop Gene Editing, Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou University, Guangzhou 510006, China

\* Corresponding author, E-mail: minl@gzhu.edu.cn

## Abstract

Intron splicing is a fundamental regulatory step in eukaryotic gene expression. However, the dynamic coupling between splicing and the release of transcripts from chromatin remains poorly understood in plants. In this study, we employed high-depth sequencing in *Arabidopsis thaliana* to analyze four subcellular RNA fractions representing sequential stages of transcript processing: chromatin-bound non-polyadenylated RNA (CB), chromatin-bound polyadenylated RNA (CB\_pA; transcripts retained on chromatin post-termination), nuclear polyadenylated RNA (Nuc\_pA; released nuclear RNA), and total cellular polyadenylated RNA (Tot\_pA). Our findings demonstrate that splicing in *Arabidopsis* proceeds progressively throughout RNA maturation, with the extent of co-transcriptional splicing significantly modulated by intron length and proximity to the polyadenylation site. By constructing regression models, we categorized genes into “fast-release” and “slow-release” cohorts, which exhibit distinct functional signatures. Further analysis revealed a positive correlation between transcript processing maturity and cellular stability. Collectively, these results uncover a splicing-driven post-transcriptional quality control mechanism in *Arabidopsis* that coordinately regulates chromatin release and transcript steady-state levels, offering new insights into the spatiotemporal dynamics of plant gene expression.

**Keywords** *Arabidopsis thaliana*, chromatin-bound RNA, splicing, chromatin release