亮点

DNA 甲基化起始的原动力: sidRNAs

王小康, 钱伟强*

北京大学生命科学院, 北京 100871

* 联系人, E-mail: wqian@pku.edu.cn

收稿日期: 2016-02-12; 接受日期: 2016-02-17; 网络版发表日期: 2016-02-23

DNA 胞嘧啶甲基化(5mC)是一种重要的表观遗传修饰, 它对于调控基因表达, 沉默转座子和维持基因组的稳定等具有重要作用. 相对于动物中 DNA 的甲基化主要以 CG 形式存在, 植物中 DNA 的甲基化有 3 种类型: CG, CHG 和 CHH(H 代表 A, T 或 C)甲基化. 这 3 种类型甲基化修饰的从头建立都必须由 RNA 介导的 DNA 甲基化途径, 即 RdDM(RNA-directed DNA methylation)来完成(图 1). 在 RdDM 途径中, 植物特有的 RNA 聚合酶 Pol IV 首先转录出一段长度为 30~40 nt 的单链 RNA, 并在 RDR2(RNA-dependent RNA polymerase 2)的作用下变为双链 RNA, 然后 DCL3(Dicer-like 3)把双链 RNA 切割成 24-nt 的 siRNAs (small interfering RNAs). 这些依赖于 Pol IV 的 siRNAs 被称为 P4siRNAs, 它们与 AGO4(argonaute 4)并通过与植物特有的 RNA 聚合酶 Pol V 转录产生的大约长度为 500 nt 左右的非编码 RNA 配对从而招募 DRM2(domain rearranged methyltransferase 2)进行甲基化的从头建立^[1,2]. 一旦 DNA 甲基化建立以后, CG 和 CHG 这两种对称的甲基化修饰分别由 MET1(methyltransferase 1)和 CMT3(chromomethylase 3)在 DNA 复制的过程中维持. 而 CHH 这种非对称的甲基化修饰, 则由 DRM2 和 CMT2 共同来维持^[2].

在模式植物拟南芥(*Arabidopsis thaliana*)中, DCL 家族有 4 个成员组成. 除了 DCL3 主要参与 24-nt

的 siRNAs 合成之外, DCL1 主要参与 21-nt 的 microRNAs 合成; DCL2 主要参与 22-nt 的病毒 siRNAs 合成; DCL4 主要参与 21-nt 的 *trans-acting* siRNAs 合成^[3]. 和 DCL3 一样, DCL2 和 DCL4 也会参与 24-nt 的 siRNAs 合成. 尽管 24-nt 的 siRNAs 在 DNA 的甲基化建立和维持方面有着非常重要的作用, 但是, 在基因组某些位点即使有较高的 siRNAs 产生, 也不会引起该位点的 DNA 甲基化^[4,5], 而某些位点即使 siRNAs 的量大幅减少, 也不会引起该位点 DNA 甲基化的降低^[6]. 此外, 24-nt siRNAs 及 Pol V 转录产物的产生需要一定程度的初始 DNA 甲基化^[2]. 因此, siRNAs 是否真正是促使 DNA 甲基化建立的决定因素还不清楚.

近期, 清华大学戚益军团队发现了一类不依赖于 DCLs 的 siRNAs(简称 sidRNAs), 这类 sidRNAs 可能参与了 RdDM 的起始, 论文发表在 2016 年 1 月 26 日的 *Molecular Cell* 上, 并被选为当期杂志的封面文章. 在这项研究中, 戚益军教授领导其团队成员通过高通量测序技术发现大量的 siRNAs 包括 24-nt 大小的 P4siRNAs 仍然存在于 DCLs 三突变体(*dcl2-1/dcl3-1/dcl4-2*, *dcl2/3/4*)和四突变体(*dcl1-9/dcl2-1/dcl3-1/dcl4-2*)中, 并且三突变体和四突变体中 siRNAs 的量并无明显的差异. 通过进一步生物信息学分析他们在 DCLs 四突变体中共鉴定到了 14360 个 siRNAs 产生位点. 在野生型中, siRNAs 主要以

引用格式: 王小康, 钱伟强. DNA 甲基化起始的原动力: sidRNAs. 中国科学: 生命科学, 2016, 46: 343-345

Wang X K, Qian W Q. The initial triggers of DNA methylation: sidRNAs. *Sci Sin Vitae*, 2016, 46: 343-345, doi: 10.1360/N052016-00098

24-nt 的形式存在, 而在 *DCLs* 三突变体和四突变体中 siRNAs 的长度从 18-nt 到 30-nt 几乎均匀分布. 作者把这类 siRNAs 称为不依赖于 *DCLs* 的 sidRNAs. 遗传学实验证据表明, 这些 sidRNAs 的产生大都依赖于 *PolIV* 和 *RDR2*. 有意思的是, 很多情况下, 从一个位点产生的 sidRNAs 在长度上呈现阶梯状分布, 即它们都有共同的 5' 端, 但在 3' 端以一个碱基的长度差异进行梯度变化. 通过 Northern blot 和高通量测序, 他们发现 *DCLs* 突变体相对于野生型 Col-0 积累了大量长度从 20-nt 到 60-nt 的 RNAs^[7].

那么 sidRNAs 的生物学功能到底是什么呢? 他们对 Col-0 和 *dcl2/3/4* 背景下的 AGO4 结合的 RNAs 进行了高通量测序, 结果发现, 在 Col-0 里 AGO4 主要结合 24-nt 的 siRNAs, 而在 *dcl2/3/4* 中, 主要结合 21~26 nt 的 sidRNAs, 且大小成阶梯状分布. 染色质免疫共沉淀(ChIP)实验表明, AGO4 结合的 sidRNAs 可以引导 AGO4 结合到目的染色质位点. 通过位点特异性 DNA 甲基化测序和全基因组水平的 DNA 甲基化测序发现, *nprp1/dcl2/3/4* (其中 *nprp1* 是 *PolIV* 的突变体)和 *rdr2/dcl2/3/4* 相比于 *dcl2/3/4*, 引起了更显著的 CHG 和 CHH 甲基化水平的降低, 说明 sidRNA 在体内可以介导 DNA 的甲基化^[7].

Northern blot 结果表明, 虽然 AGO4 在 *dcl2/3/4* 背景下结合的大部分的 sidRNAs 长度小于 *dcl2/3/4* 中的总的 siRNAs, 但是仍有小部分长度较长的 sidRNA 被检测到. 以上结果说明, 可能 sidRNAs 先被 AGO4 结合, 然后再由一个 3'-5' 外切酶对 sidRNAs

的 3' 端进行切割处理从而产生阶梯状的长度较小的 sidRNAs. 进一步研究发现, 如果在 *dcl2/3/4* 中进一步引入 *ago4* 突变, sidRNAs 的量就会大幅减少, 且互补实验表明, AGO4 及其对单链 RNAs 的结合能力对 sidRNAs 的产生至关重要. 通过对 13 个编码可能的 3'-5' 外切酶的候选基因的筛选, 他们鉴定到 *Atlg54440* (Atrimmer1) 和 *Atlg56310* (Atrimmer2) 这两个基因可能参与 sidRNAs 的产生. 通过对其突变体的 siRNAs 进行高通量测序, 发现这两个基因的突变分别引起了 715 个和 107 个位点 sidRNAs 水平的显著下调, 其中有 41 个位点被两者共同调控^[7]. Atrimmer1 和 Atrimmer2 影响的 sidRNAs 的位点相对于总的 sidRNAs 的位点非常有限. 一个可能的原因是这两个基因在功能上存在冗余, 但更有可能在拟南芥中存在多个 trimmers.

根据以上的发现以及已有的一些报道, 作者提出了一个假说: 在一个没有任何 DNA 甲基化修饰的区域, *Pol II* 和 *RDR6* 产生的 RNAs 被 AGO4 结合并在 Trimmers 的作用下完成 3'-5' 的剪切形成成熟的 sidRNAs, 并进而招募 DRM2 对该段 DNA 区域进行甲基化. 一旦该区域被 DNA 甲基化标记, 将由经典的 RdDM 途径进行甲基化的维持(图 1).

与此同时, 中国科学院上海植物逆境中心朱健康团队发现依赖于 *Pol IV* 和/或 *Pol V* 的 RdDM 靶位点中只有 16% 的位点完全依赖于 *DCLs*. 在绝大多数位点上尽管 24-nt siRNAs 大量减少, 而 DNA 的甲基化水平并没有显著变化. 相反, DNA 的甲基化水平

图 1 sidRNA 介导的 DNA 甲基化的建立和维持

Pol II 和 *RDR6* 产生的单链和/或双链 RNAs 被 AGO4 结合并在 Trimmers 的作用下完成 3'-5' 的剪切形成成熟的 sidRNAs, 然后与 *Pol II* 转录产生的非编码 RNAs 进行配对并进而招募 DRM2 对该段 DNA 区域进行甲基化. 一旦该区域被 DNA 甲基化标记, 将由经典的 RdDM 途径进行甲基化的维持

与依赖于 Pol IV 的长度为 25~50 nt 的 RNAs 水平成正相关^[8].

总之, 以上最新研究结果表明, 不依赖于 DCLs 的 sidRNAs 不仅在 DNA 甲基化的从头起始阶段具有

重要的作用, 而且在 DNA 甲基化的维持阶段可能也具有非常重要的功能. sidRNAs 的发现和功能研究使我们在认识和理解“DNA 甲基化是如何起始的”这个问题上往前迈进了一大步.

参考文献

- 1 Law J A, Jacobsen S E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 2010, 11: 204–220
- 2 Matzke M A, Mosher R A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*, 2014, 15: 394–408
- 3 Bologna N G, Voinnet O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annu Rev Plant Biol*, 2014, 65: 473–503
- 4 Chan S W, Zhang X, Bernatavichute Y V, et al. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol*, 2006, 4: e363
- 5 Arteaga-Vazquez M, Sidorenko L, Rabanal F A, et al. RNA-mediated *trans*-communication can establish paramutation at the b1 locus in maize. *Proc Natl Acad Sci USA*, 2010, 107: 12986–12991
- 6 Zhang H, Ma Z Y, Zeng L, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad Sci USA*, 2013, 110: 8290–8295
- 7 Ye R Q, Chen Z L, Lian B, et al. A Dicer-independent route for biogenesis of siRNAs that direct DNA methylation in *Arabidopsis*. *Mol Cell*, 2015, 61: 1–14
- 8 Yang D L, Zhang G P, Tang K, et al. Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Res*, 2016, 26: 66–82