

这些非编码 RNA 基因通过转录、加工、转运和修饰等系列过程最终合成非编码 RNA。在非编码 RNA 的转录过程中,在 RNA 聚合酶作用下按照碱基互补配对原则,从 5'→3' 方向进行非对称转录。基因组中可生产非编码 RNA 的 DNA 比例目前仍不清楚,最近的转录组及微阵列研究显示,在老鼠基因组中,可能有超过 3 万个非编码 RNA。目前在 NONCODE 2.0 数据库(非编码 RNA 组科学数据库)中,非编码 RNA 基因的数量为 20 多万个,其中包括 miRNA、piRNA 和 mRNA-like 非编码 RNA 等,而在人类基因组中蛋白质编码基因的数量仅为 2.5 万~3.0 万个,可见非编码 RNA 基因的数量远大于蛋白质编码基因。

近年来,随着新的非编码 RNA,特别是 siRNA、miRNA 和 piRNA 的不断发现,研究人员洞察到了小的非编码 RNA 作为细胞活动调控因子的巨大潜力。目前,在多达 800 多个物种中发现了小的非编码 RNA,遍及细菌、古细菌和真核生物界。细菌和古细菌的基因组较小,不过依然发现了不少非编码 RNA,如大肠杆菌(*E. coli*)中有大约 50 个非编码 RNA 基因。而在真核生物中,非编码 RNA 的数量远远超出生物学家以前的想象。除了在模式生物酵母、拟南芥、线虫和果蝇中发现为数不少的非编码 RNA 外,还在小鼠和人类中发现了数量惊人的非编码转录产物。有科学家预言,非编码 RNA 在生物发育过程中有着不亚于蛋白质的重要作用。但是,现在对整个非编码 RNA 的世界了解甚少,大部分非编码 RNA 都是最近才发现的,还有很多未注释的非编码 RNA,只有很少一部分非编码 RNA 进行了功能研究。虽然对这些非编码 RNA 的功能还知之甚少,但越来越多的研究表明这些非编码 RNA 在真核细胞中发挥着重要的调控作用。这种调控作用包括染色体的动力、染色质的结构、转录、转录后加工和翻译等多个层次的调节。很多这些非编码 RNA 虽然不直接参与基因编码和蛋白质合成,但是在基因转录后调控、剪切和修饰过程中具有十分重要的功能,在很多生命活动中均起着举足轻重的作用,与疾病的发生、发展、诊断和治疗有密切的关系,因而迅速成为当今分子生物学最热门的前沿研究领域之一。目前的主要任务是发现更多的非编码 RNA 及其生物功能。

二、非编码 RNA 分类

由于受中心法则的影响,非编码 RNA 的作用被忽视,因此非编码 RNA 的发现经历了一个比较漫长的过程。目前人们只是根据它们所在位置、大小、特征及功能给予分类和命名。

根据非编码 RNA 在细胞内的生物学功能,非编码 RNA 分为看家非编码 RNA(housekeeping non-coding RNA)和调控非编码 RNA(regulatory non-coding RNA)。看家非编码 RNA 一般属于组成型表达,对细胞的生存及基本功能是必须的,包括 tRNA、rRNA、snRNA、snoRNA、tmRNA 和端粒酶 RNA 等。调控非编

因子,而且与细胞的异常表型和人类重要疾病密切相关。在许多肿瘤中可检测到特有 miRNA 或 LncRNA 基因的异常表达,或 mRNA 异常可变剪接体。一些动物病毒也编码用于逃逸宿主细胞免疫攻击的 miRNA。比较分析正常生理和疾病发生过程中的非编码 RNA 的表达及其作用,将从 RNA 调控的角度揭示疾病发生机制,并为疾病诊断和治疗提供新的基因靶点和分子标记。

6. 非编码 RNA 作用机制研究

非编码 RNA 通过顺式作用或反式作用对下游靶基因的调控作用是其参与正常生理过程和疾病发生的主要机制,但是非编码 RNA 调控了哪些靶基因、影响了哪些信号通路、其网络机制怎样、与疾病发生有何关联等都是非编码 RNA 在机制研究方面需要弄清楚的问题。

7. 非编码 RNA 基因资源与 RNA 技术及其应用

非编码 RNA 基因是新发现的遗传资源和新的生物技术制高点。例如,miRNA 和 RNA 干涉技术在干细胞维持、动物和植物品种选育及病害控制等方面有重要应用前景。miRNA 治疗干预作为人类重大疾病治疗的新技术,正得到迅速发展,miRNA 药物及靶点的前景十分广阔。

二、非编码 RNA 是生物网络的元件

在传统的分子生物学研究中,普遍认为一个基因编码一个蛋白质,一个蛋白质有一个结构,一个结构完成一种功能,也就是遵循由序列到结构再到功能这样一种思维方式。越来越多的证据表明,一个基因的单独作用往往不足以主宰一个生物学事件的发生,其发生需要一组基因的同时参与,是一组或多种蛋白质协同作用的结果,这些相互作用的蛋白质构成一个复杂的生物学网络。这种以蛋白质为中心开展的大量功能与调控研究,发现了大量关于蛋白质产生、调控和代谢的信号通路及相关网络,为人类认识生命活动的本质做出了巨大贡献。

随着大量非编码 RNA 的发现,通过 RNA 与蛋白质相互作用或 RNA 与 RNA 相互作用来实现生物功能的事例也越来越多。例如,snRNAU1、snRNAU2、snRNAU4、snRNAU5、snRNAU6 同多达 75 种蛋白质组成剪接复合物,负责 pre-mRNA 的剪接。非编码 RNA 还可以通过与靶 RNA 序列匹配来定位靶标,并进一步招募功能蛋白质来行使功能。例如,miRNA 通过不完全互补配对的方式作用于靶基因 mRNA 的 3'-UTR,转录后负性调控靶基因的表达;同时,由于 miRNA 存在自身的编码基因,转录水平上也必然受到转录因子的调控,符合类似于蛋白质编码基因的转录调控理论。因此,转录因子、miRNA 和靶基因之间组成了一个复杂的调控网络:一个转录因子可以调控多个 miRNA 分子,一个 miRNA 分子又同时受到多个转录因子的调控;同时,一个 miRNA 分子能调控多个靶基因,而一个靶基因又同时受到多个 miRNA 分子的协同调控;它们之间组成了错综复杂的调

控网络,实现了对人体生命活动以及疾病发生发展和转归过程的精细调控。因此,miRNA 等大量非编码 RNA 的加入使网络结点迅速增加,网络规模成倍扩大。更重要的是,非编码 RNA 的加入增加了各种新的相互作用,使网络的内容更加丰富。所以真正能说明生物学功能的应该是一组或多组相互作用的生物大分子形成的网络,这就是系统生物学所带来的思维上的变化和研究思路的变化。

由于非编码 RNA 具有重要的生物学功能,与蛋白质一样在生命活动过程中扮演重要角色,构成了生物网络的重要节点。因而有学者引入“双色网络”的概念描绘生物网络的复杂性,其基本原理如下:生物网络包括蛋白质和功能 RNA 两类基本物质,在网络中可以简单地用两种颜色的节点表示,故称为双色网络。“双色网络”概念的引入避免了当前部分有关生物网络的研究只把蛋白质作为网络节点的局限,强调了非编码 RNA 在生命活动过程中的重要性。一些科学家认为,成千上万非编码 RNA 分子组成了巨大的分子网络调节着细胞中的生命活动,它们与蛋白质网络相对应,同时这两类网络必然有紧密的相互作用从而构成更复杂的网络。所以未来的网络至少是双色网络才更符合生命活动的实际。同时,细胞信号转导(pathway)是一个细胞内信息流动的过程,是生物分子之间的纵向作用,不可忽视。因此,关注非编码 RNA、蛋白质与传统 pathway 的相互作用,将会使生物网络更加真实、更加完整。

三、非编码 RNA 调节的多样性和复杂性

非编码 RNA 种类繁多,除了细胞内含量丰富的 rRNA 和 tRNA 之外,还包括其他 LncRNA 和短链非编码 RNA,它们参与细胞生物学功能调节的方式也多种多样的。

对 rRNA 和 tRNA 的研究最早,其功能和作用方式也研究得比较透彻。rRNA 是细胞内含量最为丰富的 RNA,其功能是作为核糖体的重要组成成分参与蛋白质的生物合成;tRNA 具有典型的三叶草结构,其功能主要是携带氨基酸进入核糖体,在 mRNA 指导下合成蛋白质;可见 rRNA 和 tRNA 在氨基酸的翻译过程中发挥重要作用。LncRNA 一般不参与蛋白质的编码,主要通过调节转录因子的结合与装配、竞争蛋白质编码基因的转录因子、与 DNA 形成三链复合物、调节 RNA 聚合酶 II 的活性和转录干扰等多种方式调节基因的转录。短链非编码 RNA 在细胞内比较常见,主要包括 miRNA、piRNA 和 siRNA,其作用的方式也多种多样。

非编码 RNA 的复杂性不仅表现在其独特的分子特征方面,而且表现在其参与调节遗传信息由 DNA 向蛋白质传递的过程中,非编码 RNA 调节遗传信息流的方式主要有以下几种。①RNA 与 DNA 的相互作用。例如,RNA 可以修饰染色体构型,进而影响 DNA 局部区域的转录活性。②RNA 与 RNA 之间的相互作用可

编码 RNA 的研究寓于 RNA 研究之中。

(二)在蛋白质为主体的研究过程中发现看家非编码 RNA

1. rRNA 和 tRNA 是相继被最早发现的非编码 RNA

rRNA 约占 RNA 总量的 80%，它们与蛋白质结合构成核糖体的骨架，参与蛋白质的生物合成。1959 年 rRNA 被分离得到，是最早发现的一类非编码 RNA。1957 年 Hoagland 等发现一类稳定的 RNA 小分子，因不与核糖体结合而不同于 mRNA 和 rRNA，Hoagland 等对它们在合成蛋白质中转运氨基酸的功能提出了假设。Crick 比较了核酸和氨基酸的大小及形状后，认为它们不可能在空间上互补，因此预测：①存在一类分子转换器，使信息从核酸序列转换成氨基酸序列；②这种分子很可能是核酸；③它不论以何种方式进入蛋白质翻译系统的模板，都必须与模板形成氢键(配对)；④有 20 种分子转换器，每种氨基酸一个；⑤每种氨基酸必定还有一个对应的酶，催化与特定的分子转换器结合。1963 年，Ehrenstein 等用实验证明了 Hoagland 发现的分子就是 Crick 预言的分子转换器，即 tRNA，这是继 rRNA 之后发现的第二类非编码 RNA。1965 年 Holley 首次测出了酵母丙氨酰-tRNA 的一级结构。特别是 1966 年 Nirenberg 和 Khorana 等破译了 RNA 上编码合成蛋白质的遗传密码，之后的研究表明这套遗传密码在生物界具有通用性，从而认识了蛋白质翻译合成的基本过程。可见，早期有关 RNA 的研究均围绕 DNA 到蛋白质的生物合成展开。由于 tRNA 和 rRNA 均不直接编码蛋白质，但在蛋白质的翻译过程中必不可少，因此是第一类被发现的非编码 RNA。但这个时期的局限在于对 RNA 生物功能认识不足。因此 20 世纪 70 年代后，人们对生物大分子的注意力更多地集中于 DNA 和蛋白质。

2. 20 世纪 80 年代核酶的发现颠覆了人们对 RNA 和蛋白质的传统认识

1981 年 Cech 发现四膜虫大核 rRNA 前体内的转录内含子是通过自动剪接被除去的。这一反应不需要传统意义上的生物催化剂——酶，催化此反应的是一种新发现的具有酶活性的生物催化剂 RNA，他将其命名 RNA 酶(ribozyme, 核酶)。由于核酶本身不编码蛋白质，但又具有重要的生物学功能，因此，核酶是一种新发现的非编码 RNA。核酶的发现首先突破了统治生物化学超过半个世纪的一个信条——酶即是蛋白质。核酶的发现表明，RNA 既可携带遗传信息，又可作为功能分子，在生命起源过程中，可能最早出现的是 RNA。此外，核酶的发现为 RNA 的产业化指明了道路，可以用它来抑制病毒和有害基因的表达。更为重要的是，核酶的发现使人们重新认识 RNA 在遗传信息传递过程中的作用，探索 RNA 生物功能的多样性和重要性。

(3) piRNA 的发现和研究

2006 年 7 月, A. Girard、A. Aravin、S. T. Grivna、N. C. Lau 4 个独立的研究组几乎同时在果蝇、小鼠、大鼠和人等物种的生殖系细胞中发现了一类新型小分子非编码 RNA, 因为它们特异性地与 PIWI 蛋白质相互作用, 所以被命名为 PIWI 相互作用 RNA, 简称为 piRNA。经过测序鉴定, 发现 piRNA 的长度为 24~30nt, 其中 85% 为 *mili-interacting RNA*。piRNA 的生物学功能主要涉及生殖相关事件, 这不仅因为 piRNA 特异性地在生殖细胞中表达, 还因为 piRNA 途径中的重要蛋白质与配子形成事件或胚胎发育直接相关。PIWI 家族蛋白质是 piRNA 途径的核心组分, 在生殖干细胞命运决定、减数分裂、精子发生等配子形成事件中具有重要作用。2007 年, Brennecke 等通过对果蝇的 3 类 piRNA——PIWI-piRNA、AUB-piRNA 和 AGO3-piRNA 的序列分析, 推测 PIWI、AUB 和 AGO3 与 piRNA 结合形成的 piRNA 复合物(piRC)具有剪切活性, 可以切割与 piRNA 序列互补的 RNA 底物, 从而提出了 piRNA 生物发生的“乒乓模型”假说, 并得到许多学者的广泛认同。2009 年, 印度科学家 S. Sai Lakshmi 及其合作者建立了一个 piRNA 数据库, 即 piRNA Bank, 以保存新发现的 piRNA 序列。piRNA Bank 收录了包括人、小鼠和大鼠的近 2000 万个 piRNA 相关序列, 经过同源序列去除、基因组定位等筛选, 得到 10 万多个在基因组中具有唯一靶位点的 piRNA 序列。

2. LncRNA 的研究

2001 年人类基因组计划完成, 宣告了后基因组时代的开始, 也掀起了从非编码 RNA 基因角度解读遗传信息的新组成及其表达调控的高潮。2002 年 Okazaki 等在对小鼠全长互补 DNA(cDNA)文库的大规模测序过程中首次发现了一类新的转录物, 即 LncRNA。LncRNA 是一类转录本长度超过 200nt 的功能性 RNA 分子, 它们缺乏编码蛋白质的能力, 位于细胞核或胞质内, 以 RNA 形式在多种层面上(如表观遗传学、转录调控及转录后调控等)调控基因的表达水平。相对于蛋白质编码序列及小分子 RNA 的研究来说, LncRNA 的研究还仅仅处于起步阶段, 随着对 LncRNA 在哺乳动物进化及疾病发生发展中作用的关注, LncRNA 的机制已成为现代遗传学研究的热点问题。2003 年就有研究表明, 另一个 LncRNA——MALAT1(*metastasis-associated lung adenocarcinoma transcript 1*)广泛存在于人类和小鼠的正常组织中。虽然早期的非小细胞癌中 MALAT1 的表达无特异性, 但在癌症转移之后, MALAT1 的表达水平较未转移者高, 说明 MALAT1 在肿瘤转移中具有提示的价值, MALAT1 的高表达与患者的低生存率密切相关。2008 年 Yu 等的研究表明, 包括 *p15* 在内的许多因表观遗传机制而发生基因沉默的抑癌基因都可转录相应的 LncRNA, 从而诱导相关疾病的发生。然而, 迄今为止, 人们对哺乳动物细胞内 LncRNA 调控机制的认识仍存在许多障碍, 包括它们的数量比例及它们的作用机制。因此提高对 LncRNA 等非编码转录本的关注, 对真核细胞基因编程中进化更新和功能性 RNA 家族的扩大是很重要的。目前发现的参与

哺乳动物基因活动的 LncRNA 已有上万个,其调控基因表达的机制存在共性。一般来说,LncRNA 主要从表观遗传学、转录调控及转录后调控 3 个层面实现对基因表达的调控。

3. 竞争性内源性 RNA(ceRNA)的发现和研究

2007 年,麻省理工学院的 Phillip Sharp 及同事开发出一种长期抑制 miRNA 基因的高效分子,称之为 miRNA 海绵(miRNA sponge),意思是吸满了 miRNA,这样 miRNA 就无法与天然靶点结合。2009 年,Seitz 提出了伪靶标假说(pseudotarget hypothesis),他认为通过生物信息学确定的 miRNA 的靶基因绝大部分是 miRNA 的竞争性抑制剂——伪靶标(pseudotarget),这些伪靶标通过竞争性结合相应的 miRNA,从而阻断其与真正靶基因的结合,因此绝大多数 miRNA 的靶基因是其相应 miRNA 的调控者,而 miRNA 只是其靶基因的被动调控对象。2010 年,Ebert 等提出天然 miRNA 海绵假说(natural miRNA sponge hypothesis),该假说认为生物体内存在对 miRNA 起负性调控作用的内源性 miRNA 海绵,在 miRNA 海绵参与的调控机制中存在一个 miRNA 丰度阈值,只有相应 miRNA 的丰度在该阈值以上能够抑制其关键靶基因的表达时该 miRNA 海绵才能发挥调控作用。基于上述一系列关于 RNA 转录体之间相互对话的研究,2011 年,Salmena 等提出了竞争性内源性 RNA 假说(competing endogenous RNA hypothesis,ceRNA 假说),该假说认为 ceRNA 能与其他 RNA 转录体竞争相同的 miRNA,形成大规模的转录调控网络,从而扩大人类基因组的遗传信息。ceRNA 假说本质上是“RNA→miRNA”的作用方式,是对传统的“miRNA→RNA”学说的补充。ceRNA 通过与 miRNA 反应元件(MRE)竞争相同的 miRNA,从而调节转录体的表达,不管是否编码蛋白质,对于富余的 RNA 转录体来说,彼此都能竞争与 miRNA 的结合,从而构成庞大的 ceRNA 网络(ceRNET)。因此,ceRNA 是一种内源性 RNA 转录体,能与其他 RNA 转录体竞争相同的 miRNA,从而实现相互间的交流与调节。基于上述理论,ceRNA 在功能层面上是 miRNA 的拮抗子,与 ceRNA 基因本身是否编码蛋白质无关,蛋白质编码基因、假基因(pseudogene)、LncRNA 及环状 RNA(circular RNA)均可能成为理想的 ceRNA。其中假基因是最配对的 ceRNA,它们与对应基因几乎拥有完全一致的 MRE,能够非常特异性地拮抗 miRNA 与基因的 3'-UTR 的结合而达到调节 miRNA 功能的效果,是一类理想的 ceRNA。

二、非编码 RNA 研究过程中的标志性成果

非编码 RNA 的研究贯穿了整个核酸研究的过程,推动了生命科学研究的发展。下面以诺贝尔奖获取情况为例介绍非编码 RNA 或 RNA 研究过程中产生的一系列标志性成果。

(1)德国科学家科塞尔(A. Kossel)发现组成核酸的 4 种碱基,因为他在核酸化

是这三类小非编码 RNA 中长度最长的非编码 RNA, 它的生物合成与 PIWI 亚家族蛋白密切相关。非常有趣的是, piRNA 的生物合成不依赖于 Dicer。piRNA 主要通过异染色质形成或通过 RNA 去稳定性作用参与转座子的沉默而发挥抑制作用。内源性 siRNA 的生物合成与 miRNA 非常相似, 也依赖于 AGO 家族蛋白。与 miRNA 不同的是, 内源性 siRNA 主要来源于长的 dsRNA, 且依赖于 Dicer 内切核酸酶的作用, 却不需要 Drosha 内切核酸酶的作用。它们比 miRNA 更短, 只有 21nt 左右。有一小部分内源性 siRNA 也被发现可以作为转录后调控子靶向抑制 RNA。本书将详细介绍这 3 类真核细胞小非编码 RNA 的生物合成过程。

一、miRNA 的生物合成

(一) 经典的 miRNA 的生物合成

截至目前, miRBase ver. 18(<http://www.miRbase.org>) 数据库已经记录并注释了不同物种基因组的 miRNA, 包括人类的 miRNA 2652 个、秀丽隐杆线虫的 miRNA 292 个、果蝇的 miRNA 252 个、斑马鱼的 miRNA 346 个、鸡的 miRNA 734 个和拟南芥的 miRNA 481 个。人类基因组中超过 1/3 的基因均受到 miRNA 的调控, miRNA 已经成为基因组中最大的一类基因调控分子。而且, 许多动物的 miRNA 在各个系统发育中都是非常保守的, 如 55% 的秀丽隐杆线虫的 miRNA 和人类的 miRNA 具有高度的同源性, 这表明 miRNA 在整个动物的进化中发挥了重要的作用。经典的 miRNA 生物合成途径主要包括, 在细胞核中转录, 经 pri-miRNA 加工生成 pre-miRNA 出核进入细胞浆, 在细胞浆中进行广泛加工, 最后形成成熟的 miRNA。成熟的 miRNA 锚定 AGO 蛋白形成 miRNA 诱导的沉默复合物, 最后发挥 miRNA 的靶向抑制效应, 具体步骤如下所述。

1. miRNA 基因及其转录

大约 60% 的 miRNA 位于基因间隔区 (intergenic region), 其余 40% 则位于蛋白质编码基因或其他转录元件的内含子上。很多基因间 miRNA 成簇分布, 这样的 miRNA 有可能是在共同的启动子调控下表达的, 因为它们具有相似的表达谱。与上述情况相反, 内含子 miRNA 则更可能是与其所在基因的启动子共表达, 因为大多数内含子 miRNA 的表达与寄主 mRNA 的表达相似。研究表明, 大约有 1/3 的已知 miRNA 嵌入在蛋白质编码基因的内含子中, 与宿主基因共转录, 从而协同调控 miRNA 和蛋白质的表达。在某些情况下, 内含子 miRNA 与基因编码蛋白质一起参与调控生物过程。例如, miR-33 家族成员可与其嵌入的胆固醇调节元件结合蛋白 (SREBP) 基因协同作用减少胆固醇流出, 促进胆固醇合成。但是, 也有些内含子 miRNA 具有独立的启动子。

50% 以上的哺乳动物 miRNA 在基因组中均以成簇的形式存在, 虽然这些成簇的 miRNA 中也有个别 miRNA 具有独立的基因启动子, 但是大多数成簇的

的 RNA 片段都有 3' 端两个碱基的突出和 5' 端的磷酸基团。由于 Dicer 剪切位点由之前的 Drosha 剪切位点决定, 因此可以说 Drosha 通过间接的方式决定了成熟 miRNA 的最终序列。

在动物细胞核中, Drosha 的切割产生了成熟 miRNA 的一端(3' 端)。另一端的切割成熟是由细胞质中的 Dicer 酶切割产生的。Dicer 首先是在研究 siRNA 引起的基因沉默中被发现的, 后来发现它在 miRNA 的成熟过程中也起着重要的作用。Dicer 参与 miRNA 的成熟过程与 RNA 干扰中产生双链 RNA 的过程很相似: 首先, Dicer 识别 pre-miRNA 的双链部分, 其与茎环基部的 5' 端被磷酸化、3' 端有突出的结构有很高的亲和力; 然后, 茎环基部的两圈螺旋解开, Dicer 对两条链都进行切割, 将前体的其余部分切掉, 生成 5' 端磷酸化、3' 端有 2nt 突出的不完全配对的双链。

Dicer 在所有的真核细胞, 包括粟酒裂殖酵母、植物和动物中具有高度的保守性。有些组织含有多种 Dicer 的同源物, 而不同亚型的 Dicer 也具有不同的作用。例如, 果蝇的 Dicer1 是果蝇 miRNA 合成必需的, 而 Dicer2 则在 siRNA 的产物形成中发挥作用。

Dicer 主要与 dsRNA 结合蛋白结合发挥作用。例如, 在果蝇的 pre-miRNA 加工过程中, Dicer1 与 Loquacious(LOQS, 也称为 R3D1) 蛋白结合, 该蛋白质含有 3 个 dsRNA-结合结构域(dsRBD)。人类的 Dicer 主要与两个高度同源的蛋白质 TRBP(transactivating response RNA-binding protein, 也称为 TARBP2) 和 PACT(protein activator of PKR, 也称为 PRKRA) 相互作用发挥剪切作用。TRBP 和 PACT 本身不具有活性, 它们主要参与形成 RNA 诱导性沉默复合物(RNA-induced silencing complex, RISC)。

5. AGO 的加载

完全加工的 miRNA 与 RISC 中的 AGO 家族蛋白偶联, 成为它们的特异性决定子。在人类, Dicer、TRBP(和/或 PACT) 和 AGO 蛋白共同形成 RISC 加载复合物(RISC loading complex, RLC), 在 RISC 的组装中发挥作用(果蝇中, RLC 包含 Dicer1、LOQS 和 AGO1)。尽管目前关于 RLC 是如何与 RNA 结合并有利于 AGO 的加载的确切机制还不清楚, 但是目前的证据已经表明, 经 Dicer 剪切释放的 miRNA 双链的一端可以与 RLC 中的 TRBP 结合, 而另一端则与 AGO 蛋白发生相互作用。有关 miRNA 链的选择以及 RISC 组装的确切机制在果蝇中研究的已经比较透彻。R2D2, 一个含有 2 个 dsRBD 的蛋白质, 与 Dicer2 形成稳定的异染色质复合后, 再结合到 RNA 双链中较为稳定的那条链上, 引导 AGO2 作用到 RNA 双链上。AGO 蛋白具有内源性的核酸酶活性(沉默子活性), 可以降解 siRNA 双链和一些 miRNA 双链上的亲链(parent strand)。与 siRNA 双链不同的是, 大多数的 miRNA 双链均在中间含有不匹配的区域, 有些 AGO 蛋白(如人类的 AGO1、AGO3 和 AGO4; 也分别称为 EIF2C1、EIF2C3 和 EIF2C4) 缺乏沉默子的活性, 可以保护亲链以免其被剪切。RNA 解旋酶活性被认为可以介导 miRNA 双

chinim)及其他基因(如 *Spindle-E*、*Krimper* 和 *Maelstrom*)的突变可以导致果蝇卵巢中 piRNA 的缺失,表明这些基因参与了果蝇 piRNA 的生物合成。然而,有关这些基因编码蛋白质的确切功能作用仍不清楚。

在果蝇发育过程中,PIWI 是以“乒乓”循环方式实现 PIWI 的持续生物合成。AUB 和 PIWI 可以通过母系遗传到子代中,同样的现象在鱼的胚胎中也被发现。最近的研究表明,通过母系遗传到胚胎中的 piRNA 以表观遗传调控的方式沉默转座子的转录。这种现象也表明,有些遗传性的发育不全可能与 piRNA 的遗传丢失密切相关。与 PIWI 蛋白相关的 piRNA 的母系遗传可能作为“种子”在果蝇的胚胎发育中启动了 piRNA 的“乒乓”循环的生物合成过程。因此,这种“乒乓”循环的生物合成过程在亲代和子代之间是可以传递的。

(三)小鼠中 piRNA 的生物合成

piRNA 在染色体上的分布极不均匀,在小鼠中它们主要分布于 17 号、5 号、4 号、2 号染色体上,很少分布于 1 号、3 号、16 号、19 号和 X 染色体上,基本不分布于 Y 染色体上。在哺乳动物中,已经鉴定了两类 piRNA。一类是粗线前期 piRNA,在减数分裂粗线前期表达,主要来源于富含重复序列和转座子的 piRNA 簇。这类 piRNA 的生物合成与果蝇相似,粗线前期 piRNA 的生物合成主要与鼠 PIWI 家族蛋白质 MILI 和 MIWI2 相关。另一类 piRNA 主要在减数分裂粗线期表达丰富,其生物合成主要与 MILI 和 MIWI 相关。粗线期 piRNA 在精母细胞中表达非常丰富:大于 80 000 物种的粗线期 piRNA 均来自于超过 200kb 的基因组 piRNA 簇。这些 piRNA 簇表现出明显的链不对称性,表明这类 piRNA 是从一个或少数几个粗大的转录本加工而来。研究发现,鼠的 piRNA 生物合成机制与果蝇的一样,其粗线前期 piRNA 的合成机制也是“乒乓”循环机制(图 2-5)。然而,胚胎、新生鼠和成鼠的 piRNA 的研究表明,鼠的 piRNA 的“乒乓”合成机制并不伴随整个雄性生殖系的发育过程。此外,粗线期 piRNA 也不具备“乒乓”生物合成的特点。尽管果蝇和小鼠粗线前期 piRNA 的生物合成机制有所相似,但二者还是有所区别。在小鼠的睾丸中,特异性反转录转座子的启动子区呈甲基化状态。MILI 和 MIWI2 的突变可以去除长散布核元件(long interspersed nuclear element 1, LINE1)和脑池内 A 颗粒(intracisternal A particle, IAP)的甲基化,可导致雄性鼠患不孕不育症。MILI 和 MIWI2 在胎儿雄性生殖细胞转座子的 DNA 从头甲基化中具有重要作用。综上所述,鼠的 piRNA 通过 DNA 甲基化修饰对靶基因实现转录沉默。

文献表明,鼠的 piRNA 在其 3'端具有 2'-O-甲基化,与植物 miRNA 的结构特点相似,但不同于哺乳动物的 miRNA。植物 miRNA 2'-O-甲基转移酶 HEN1 在小鼠中的同源蛋白 mHEN1 在体外培养的睾丸中特异性表达,并使 piRNA 的 3'端发生甲基化,而 HEN1 在果蝇中的同源蛋白 PiMET 则介导了 piRNA 3'端的 2'-O-

转录而成。在果蝇的 siRNA 生物合成中,内源性 siRNA 可以与 AGO2 特异性结合。小鼠的小 RNA 表达谱研究表明,内源性 siRNA 在卵母细胞中广泛存在,而在胚胎细胞中较少发现。鼠的内源性 siRNA 的生物合成也依赖于 AGO 家族蛋白(鼠的 AGO2,也称为 EIF2C2)。研究表明,植物和秀丽隐杆线虫中也存在内源性 siRNA,但是,植物和秀丽隐杆线虫中的内源性 siRNA 的生物合成要比果蝇和哺乳动物复杂得多。在植物和秀丽隐杆线虫中,内源性 siRNA 主要依赖于 RNA 依赖的 RNA 多聚酶(RNA-dependent RNA polymerase, RdRP)产生。而果蝇和哺乳动物中缺少 RdRP,内源性 siRNA 是不依赖于 RdRP 合成的一个新的内源性 siRNA 的类型。

(一)果蝇中 Dicer2 产生的内源性 siRNA 的生物合成

与外源性 siRNA 相似,但是不同于 miRNA,内源性 siRNA 依赖于 Dicer2,而非 Dicer1。与外源性 siRNA 不同的是,外源性 siRNA 的生物合成需要 R2D2,而内源性 siRNA,尤其是那些来源于长链茎环结构的 siRNA 则主要依赖于在果蝇 miRNA 合成通路中发挥重要作用的 loQS。目前的研究只知道果蝇中的 Dicer2 与 loQS 结合,但是 loQS 为什么选择 Dicer2 而不选择 Dicer1 参与内源性 siRNA 的合成,以及 loQS 如何与 Dicer2 发生作用还不清楚。另外,Dicer2-loQS 复合物如何识别它们的作用底物决定其剪切位点的机制也不清楚,如果能将这些问题阐明,将对解析内源性 siRNA 的生物合成机制具有重要的意义。

内源性 siRNA 前体的第一种来源是转座子的互补正反义链。这两条链既可以来源于蛋白质编码基因的转录,也可以来源于基因组中未注释的区域。这些转录本可以不来源于同一转录位点,因此,dsRNA 前体的两条链含有许多自然的错配序列和凸起。内源性 siRNA 前体的第二种来源是那些单链,但可以自我杂交的转录本,这些转录本可以形成长的茎环状结构,这些茎环结构与 miRNA 前体茎环结构的区别在于形成内源性 siRNA 茎环结构的颈更长。内源性 siRNA 前体的结构特点很可能决定了 dsRNA 被加工因子 Dicer2 和 loQS 识别。长 dsRNA 的加工需要 Dicer2,并需要与 AGO2 相互作用。loQS 很可能在 Dicer2 与错配的 dsRNA 结合的过程中发挥重要作用,并对 RISC 复合物的组装有贡献。有些 piRNA 产生的前体也是内源性 siRNA 的来源。像 piRNA 一样,果蝇的内源性 siRNAs 的 3'端也具有 2'-甲基基团,可以被 HEN1 甲基转移酶甲基化。piRNA 和内源性 siRNA 的区别在于:①piRNA 长 24~29nt,而内源性 siRNA 长 21nt 左右;②参与 piRNA 和内源性 siRNA 生物学合成的蛋白质分别为 PIWI 蛋白和 AGO2;③piRNA 主要在精细胞中表达,而内源性 siRNA 则广泛存在于各种细胞;④piRNA 来源于 ssRNA 而 piRNA 则主要来源于 dsRNA。反转录转座子既可以转录成 piRNA,也可以转录成内源性 siRNA,然而正反义链的数量可能是不对等的,这会导致 dsRNA 和 ssRNA 的混合存在。不管进一步如何编辑,dsRNA 都会从细胞核进入细胞浆,作

列(人的重复序列为 TTAGGG)构成的特殊结构,通过防止染色体末端 DNA 被识别为 DNA 损伤位点而发挥保护染色体的功能,从而维持基因组的完整性和稳定性。端粒的缩短与染色体的稳定性和致癌过程密切相关。端粒长度异常被认为是恶性肿瘤染色体变化的早期征兆,其功能异常或缺陷将导致染色体不稳定、降低 DNA 的修复能力,导致复杂的细胞遗传学异常,并出现“断裂-融合-桥循环”(breakage-fusion-bridge cycle)等现象。“断裂-融合-桥循环”是一种染色体畸变过程,它从染色单体断裂开始,往往最后造成一个染色体缺失而另一个染色体重复。

长期以来,端粒区域被认为是转录沉默的区域,2007 年 10 月, Azzalin 等在 *Science* 杂志上首次报道了端粒及亚端粒区域可以转录一种含有端粒重复序列(UUAGGG)的非编码 RNA——端粒 RNA(telomeric repeat-containing RNA, TERRA),打破了该区域不具有转录特性的观点。端粒 RNA 的转录起始位点位于亚端粒区,该分子从亚端粒区域向染色体末端方向转录,转录产物仅定位于细胞核内。端粒 RNA 的长度有较大变化,为 100bp~9kb。端粒 RNA 具有如下生物学功能:①端粒 RNA 参与端粒结构的形成;②端粒 RNA 通过抑制端粒酶活性参与端粒长度的调控;③端粒 RNA 参与细胞分化及发育;④端粒 RNA 导致端粒丢失和 DNA 损伤。此外,端粒 RNA 可能参与了肿瘤的发生发展。在人类喉癌、结肠癌和 B 细胞淋巴瘤组织中,其端粒 RNA 水平比相应正常组织的端粒 RNA 水平显著降低,并且随着肿瘤恶性程度的升高,端粒 RNA 水平逐渐降低,提示端粒 RNA 极有可能参与肿瘤的发生和进展。

LncRNA 参与了染色体端粒长度的维持和稳定,在细胞核、末端着丝粒和亚着丝粒的异染色质结构中起作用。通过高通量的 tiling 芯片技术,在疟原虫中筛查出了 60 个假定的疟原虫 LncRNA,通过对其 GC 含量、保守性、表达谱及与邻近基因表达相关性的研究,发现了一个定位于 4 号染色体、编码端粒相关重复原件(TARE)的长链非编码 RNA 基因,命名为 *LncRNA-TARE-4L*。通过对 *LncRNA-TARE-4L* 位点的进一步研究,发现 *LncRNA-TARE-4L* 以家族形式存在,在 28 个恶性疟原虫染色体末端,有 22 个具有同源性的 *LncRNA-TARE* 位点。通过对 22 个端粒相关 LncRNA 的研究,发现 *LncRNA-TARE* 在所有检测的染色体末端存在协调、阶段特异性表达,并且含有两个长度大约为 1.5kb 和 3.1kb 的转录本。*var* 毒力基因与预测的 *LncRNA-TARE* 基因毗邻,且 *LncRNA-TARE* 序列在 Ups-B 型 *var* 基因转录因子结合位点处高度富集。

拟南芥中的一些端粒转录子可以形成 siRNA 并促进端粒 CCCTAAA 复制子的甲基化。虽然在拟南芥和啤酒酵母细胞中发现 *LncRNA-LERRA* 来源于端粒,但是事实上大部分来源于着丝粒中端粒 DNA 的残留。在哺乳动物中只发现了 UUAGGG 端粒转录子,但是在拟南芥中端粒可以被转录,从而暗示在某些环境下端粒可以成为启动子。通过 RNA 异染色质免疫共沉淀法发现 *LncRNA-LERRA* 在异染色质化中发挥作用。研究发现,在单细胞原生生物中端粒非编码 RNA 在

酸化的非编码 RNA,提示这些非编码 RNA 负责重组热点,但是具体机制仍然不太清楚。非编码 RNA 调控染色体重排存在多种机制,有一种假说认为:在真核基因组中,一些非编码 RNA 能够影响复制时间从而影响染色体重排。延迟复制时间(DRT)和延迟有丝分裂染色体浓缩(DMC)通常会在肿瘤和受到电离辐射的细胞中发现,并且表现出永久的染色体不稳定性。研究证实,LncRNA 在 DRT 中发挥重要作用,而且含有 DRT 染色体的细胞也会发生更高概率(30~80 倍)的染色体重排,可见特异的 LncRNA 转录稳定性直接关系到染色体的重排。

非编码 RNA 的一个共性就是这些转录子能够招募蛋白质聚集在一个位点、区域或整条染色体上。一些非编码 RNA 只在促使形成 DSB 中发挥作用。典型的 DNA 复制和转录是独立的,但是,停滞的复制或增加的异常转录等因素可能导致复制叉崩溃及复制与转录之间的冲突。同样,双向转录中两个 RNAP 复合体在移向对方时也可能发生冲突。冲突的一种结果就是产生大、稳定甚至双重的 R 环。这些 R 环通常会以定向的脱氨基作用形成 SSB 或 DSB,从而导致染色体重排。非编码 RNA 转录子是亚着丝粒和端粒区域 RNA-DNA 杂交信息的潜在促进因子。LncRNA-TERRA 转录子会在 CpG 岛甲基化水平降低患者的端粒处累积。这些患者端粒功能的丧失可能是由于复制叉崩溃频率增加,更深层次的原因是由于端粒 RNA 也能够抑制端粒酶的活性从而阻止复制叉的修复。酵母中的 tDNA、长末端复制子(LTR)、Ty 成分或反向重复序列都能导致复制叉停滞从而诱发 DSB。在直接复制裂殖酵母的 LTR 时,为了协调后随链的合成和阻止单链退火,LTR 会招募一个 DNA 结合因子、开关激活蛋白 1(Sap1)从而稳定基因组在复制叉停滞过程中的持久性。通过 LTR、Sap1 突变子不仅能表现复制叉的直接进展,而且能表现染色体重排对反转录转座子的危害。通过 rDNA 复制子可以阻止复制叉的进程,Sap1 的作用就是协同 rDNA 达到定向复制和阻止这些复制子的有丝分裂重组。此外,反向重复序列(包括多种类型的非编码 RNA)能够产生双向转录。总的来说,LTR、tRNA 和反向重复序列是划分哺乳动物及酵母的脆性位点如何通过一系列过程导致高频率 DSB 的特征。

有学者提出,脆性位点整合在核区域并由 p53 结合蛋白 1(53BP1)和其他染色质相关蛋白标记。一些 DNA 损伤可以通过 DNA 修复机制得到修复,而且核小体的屏障作用也能进一步降低基因不稳定性和潜在损伤的危害。应该还有其他机制用于解释通过非编码 RNA 基因区域、启动子等的缓慢移动而导致调控网络的变化。因此,一个简单重排能够导致基因网络的显著变化,从而为组织(或种群)提供选择性优势和引导重排的保留。因而脆性位点的存在是十分有意义的。

第二节 非编码 RNA 在转录水平对基因表达的调节

无论真核生物还是原核生物,转录水平的调节都是基因表达调控最重要的环

细胞质内定位、RNA 编辑和 mRNA 的稳定性等多个重要环节。许多类型的非编码 RNA 在转录后水平参与了基因表达的调节。

一、LncRNA 在转录后水平对基因表达的调节

LncRNA 在转录后水平可以通过与互补的 mRNA 形成 RNA 双链体——dsRNA,掩盖 mRNA 内部与反式作用因子结合必需的主要元件,影响 mRNA 的加工、剪接、转运、翻译和降解等过程,从而正性或负性调节基因的表达。

上皮细胞的 Zeb2-mRNA 具有非常长的 5'-UTR,UTR 的一个内含子中含有与核糖体结合并有效启动翻译的位点,而该内含子的保留则依赖于该区与其互补反义转录物(Zeb2 antisense RNA)的结合,并掩蔽剪切位点,从而抑制该内含子的剪切,促进该位点与核糖体的结合,提高 Zeb2 蛋白的表达量。哺乳动物 A 型利尿钠肽前体(natriuretic peptide precursor type A,NPPA)的基因位点也存在天然反义转录物,影响 NPPA mRNA 内含子的剪切。推测 NPPA 的反义链可能也通过与 NPPA mRNA 形成 dsRNA,影响 NPPA mRNA 的剪接过程。在人类和小鼠中枢神经系统中,RNA 聚合酶 III 可以转录 BC1 和 BC200 两种非编码 RNA,其中 BC1 与很多神经系统特异性 mRNA 互补,并抑制靶标 mRNA 的翻译功能。另外,酵母 KCS1 非编码 RNA 推测也是通过与靶标 mRNA 形成 dsRNA,抑制目标 mRNA 的翻译功能,从而参与酵母的磷胁迫反应。

核质转运是真核细胞基本的生命活动之一,活化 T 细胞核因子(nuclear factor of activation T cell,NFAT)可以与很多蛋白质相互作用,其中包括核质运载体 b(importin-b)输入蛋白家族。研究发现,一种 LncRNA 可以通过调节 NFAT 来调控物质的核质转运。信号肽识别颗粒 SRP 是由蛋白质和 LncRNA 组成的复合体,最近发现 SRP-LncRNA 像一个可调控的开关,可以激活 SRP 与受体的结合,在蛋白质识别和转运过程中发挥重要作用。在一些肿瘤中发现的 LncRNA MALAT1 参与募集 pre-mRNA 剪接因子至细胞核内的基因转录位点,调节 pre-mRNA 剪接。

此外,LncRNA 可作为小 RNA 的前体发挥作用。这些 LncRNA 可在 Dicer 酶切割下产生 miRNA,或经 Drosha 酶剪切并在 PIWI 蛋白参与下加工成 piRNA。此外,还可经 Dicer 酶或 Drosha 酶切割加工成很多功能和生物发生机制尚不明确的小 RNA。因此,LncRNA 可充当 miRNA、piRNA 或其他小 RNA 前体而发挥作用。例如,成熟 miR-21 有一个全长为 3433bp 的初级转录产物 pri-miR-21,它包含完整的 5' 端帽子结构和 3' 端 poly(A)尾巴,但是没有明确的 ORF 框架,说明 pri-miR-21 一方面可以充当 miR-21 的前体,另一方面也充当一个 LncRNA。另外,Xist 和 Tsix 是两个调节哺乳动物 X 染色体失活的 LncRNA,有证据表明它们也被加工产生小 RNA。因此,这些 LncRNA 经过剪切、加工成成熟的小 RNA 后

酰化酶。Xist 并不足以维持体细胞杂交体中 X 染色体的失活,它也不是维持 X 染色体失活所必需的,在维持 X 染色体失活中起主要作用的是 DNA 的甲基化。

近年来的研究表明,在 *Xist* 基因的内部存在一个 1.6kb 的 ncRNA-RepA,它是 Polycomb 复合物的一个组分,该复合物能直接作用于靶基因 *PRC2*。在 RepA 的作用下,*PRC2* 首先被招募到 X 染色体上,与 Ezh2 一起起 RNA 结合亚单位的作用。反义 Xist RNA 能够阻止这种作用。RepA 的消耗也能废除全长 Xist 的作用以及 X 染色体上组蛋白 H3 上第 27 位赖氨酸(H3K27)的甲基化。同样,*PRC2* 不足能够抵偿 Xist 的作用。因此,RepA 和 *PRC2* 对于 X 染色体失活的传递和启动是必需的,是辅助 Polycomb 复合物发挥功能的辅助因子。

2. LncRNA 参与基因组印记

基因组印记通常为共价或非共价标记的 DNA 甲基化,还包括组蛋白乙酰化、甲基化等修饰,使带有印记修饰的子代等位基因呈现不同的转录活性,最终导致其中一方在功能表达上保持沉默。虽然印记基因只占人类基因组的不足 5%,但在胚胎、胎儿的生长和出生后的发育中却起着至关重要的调节作用,对行为和大脑的功能也有很大的影响。印记基因通常以基因簇形式存在,形成印记区,印记区基因转录活性受差异性甲基化区域调控。

迄今为止,在人和小鼠等哺乳动物的基因组中已发现了数十个基因印记区域,在这些区域中发现了很多 miRNA、C/D box snoRNA、反义链 RNA 和其他大非编码 RNA。一个印记区内至少存在一个非编码 RNA 基因,其中 *LncRNA* 基因出现频率高,提示 *LncRNA* 基因转录受基因组印记调控。*H19* 是第一个发现的与基因组印记密切相关的基因,其经过剪接及多聚腺苷酸化后输送至胞浆内,并可累积达到较高浓度。人类 *H19* 基因位于染色体 11p15.5,紧邻 *IGF2* 基因,人体内 *IGF2* 基因和 *H19* 基因分别表现出父本和母本互补表达模式。印记区 *H19* 与胰岛素样生长因子 2(insulin-like growth factor 2, *IGF2*) 反义,是 miR-675 前体,呈母源表达。在发育过程中,*H19* RNA 在胎盘、胚胎及大部分胎儿组织中高表达,但出生后表达水平明显降低。不同哺乳动物中 *H19* RNA 具有明显的序列相似性,并且拥有一个相似的二级结构。*H19* 基因调节 *IGF2* 的印记表达主要通过绝缘子介导对增强子的竞争过程来实现的。在母本染色体上由增强子和绝缘子所构成的调控区域位于 *H19* 基因上游的 2~4kb 区域,绝缘子序列没有甲基化,因此,可顺利结合 CTCF 因子,从而阻止增强子对基因的增强作用,同时活化 *H19* 基因。反之,在父本染色体上,绝缘子序列被甲基化而不能结合 CTCF 因子,故允许增强子对 *IGF2* 发挥活化作用,同时 *H19* RNA 的表达被抑制。绝缘子序列如同一个“印记控制中心”,控制 *H19* 和 *IGF2* 的交替表达。*H19* RNA 和 *IGF2* 因子互补表达模式显示二者在发育过程中具有特殊作用,*IGF2* 是一种刺激细胞再生的生长因子,而 *H19* RNA 具有促进细胞分化的作用。如果改变 *H19* 基因上游绝缘子序列的甲基化状态,将导致 *H19* 和 *IGF2* 基因印记效应丢失而使二者双等位表达,结果导致细胞的

恶性增生。

基因组印记除了与 *H19* 基因簇有关外,还需要 *Kcnqlot1*、*Air* 及 *Nespas* 基因的参与,它们通常是父源表达的,通过卵母细胞中启动子的 DNA 甲基化抑制母系等位基因的表达。*Kcnqlot1* 和 *Air* 启动子的缺失实验结果表明,LncRNA 或其转录产物分别对 *Kcnqlot1* 和 *lgf2r/Air* 印记基因簇的沉默是必需的,而且对长达 800bp 的印记基因簇抑制性组蛋白标记(H3K27 三甲基化和 H3K9 双甲基化)以及某些基因的 DNA 甲基化也是必需的。

3. LncRNA 参与基因的甲基化修饰

甲基化修饰是基因表达失活的重要表观遗传学机制。越来越多的实验结果表明,LncRNA 参与了基因的甲基化修饰。*HOX* 基因的反义基因间 RNA (*HOX antisense intergenic RNA*, *HOTAIR*) 是一个长度为 2.2kb 的基因,定位在哺乳动物 12q13.13 的上 *HOXC* 位点,不编码任何蛋白质。这个 LncRNA 在原发性和转移性乳腺癌中显著上调,高出正常乳腺组织的 2000 倍,而且 *HOTAIR* 的表达与肿瘤转移和预后差有关。同样,与正常结肠组织比,*HOTAIR* 在结肠癌组织中的表达量增加,而且高表达 *HOTAIR* 的患者预后差。*HOTAIR* 在肝癌病例中也显著高表达,而且可以预测肝移植后肝癌的复发。有研究表明,*HOTAIR* RNA 与哺乳动物多梳蛋白抑制性复合物 2 (*PRC2*) 密切相关,多梳蛋白群介导发育过程中控制分化的几千个基因的转录抑制,而且在干细胞的多能性及人类癌症中具有一定作用。目前,尽管 *HOTAIR* 重编程染色体状态而启动癌症转移的表型非常清楚,但是 *HOTAIR* 的作用机制仍有待阐明。

母系表达基因 3 (*maternally expressed gene 3*, *MEG3*) 是第一个被发现有肿瘤抑制功能的 LncRNA,已发现 12 个 *MEG3* 亚型。*MEG3* 表达于多种正常组织中,特别在脑组织中呈现高表达,但在脑癌及一些肿瘤细胞株中不表达,且 *MEG3* RNA 的异常表达可抑制癌细胞的生长。*MEG3* 能与 cAMP、p53、鼠双微基因 2 (*MDM2*) 和生长分化因子 15 相互作用,*MEG3* 作为调控性的 RNA 可通过依赖 p53 和非依赖 p53 起作用。*MEG3* 本身的表达受表观遗传学的控制,在多种癌症类型中 *MEG3* 存在异常 CpG 甲基化。通过对多发性骨髓瘤、急性髓性白血病、骨髓增生异常综合征病例 *MEG3* 基因启动子甲基化状态分析发现,在 57% 的多发性骨髓瘤中存在异常的甲基化谱,并且甲基化谱与疾病的亚型和阶段相关,同时发现,在 34.9% 的 MDS 和 47.6% 的急性骨髓性白血病中有异常甲基化,但在甲基化状态与核型、疾病亚型和预后评估系统间没有重要联系。但采用 Kaplan-Meier 生存评估方法,发现 *MEG3* 启动子异常高甲基化的急性骨髓性白血病、骨髓增生异常综合征患者总的生存率下降。有报道显示,与正常肝细胞比,*MEG3* 的表达在肝癌细胞中下调 210 倍,增加 *MEG3* RNA 可降低肝癌细胞生长、诱导细胞凋亡,miR-29 能调控 *MEG3* 的表达并增加 *MEG3* 的表达水平。目前,*MEG3* 在细胞生物学上的重要性及在肿瘤发生中的作用还在不断研究中。此外,其他的一些

EZH2(enhancer of zeste homolog 2)参与了 siRNA 介导的转录基因沉默(TGS)。

通常情况下,完全互补的 dsRNA 前体诱导的染色质重塑是由 AGO2 引发的,而发卡状的 dsRNA 前体诱导的染色质重塑是由 AGO1 引发的。有趣的是,AGO 在 TGS 中的作用早于靶标启动子的组蛋白甲基化,当靶标沉默态组蛋白修饰(H3K9 甲基化)增加时,AGO1 明显减少,证明 AGO1 对 H3K9 的双甲基化是必需的。EZH2 作为组蛋白甲基转移酶,在一定程度上能够引起 H3K27 的甲基化,而甲基化的 H3K27 作为锚点可募集多余的 Polycomb 蛋白家族,从而参与沉默态染色质的形成,导致 TGS。DNMT 与 EZH2 抑制的基因之间的结合依赖于 EZH2 的存在;EZH2 对于其靶向的启动子甲基化是必需的,提示 EZH2 作为募集 DNA 甲基转移酶的平台,参与了 DNA 的甲基化。总之,TGS 的建立与维持需要多种不同的蛋白质;通常 AGO1、DNMT3a 及 HDAC-1 对于起始的沉默是必需的,而 DNMT1 对于维持沉默是必需的。RNA 多聚酶 II (RNA polymerase II, RNAPII)也参与了 siRNA 介导的 TGS,研究报道 RNAPII 与 AGO1 免疫共沉淀于启动子区域,但尚不清楚 RNAPII 是如何发挥作用的。

近年来的研究发现,siRNA 除了通过序列特异性切割 mRNA 而沉默靶基因表达外,还可通过调节 DNA 甲基化及其相应组蛋白的甲基化而调控靶基因的表达。siRNA 引发 DNA 甲基化在植物中的研究较早。在拟南芥中,DNA 甲基化是一个由 RNA 和位点特异性 DNA 甲基化转移酶起始的渐进过程。它分为 siRNA 信号的产生、从头合成 DNA(de novo DNA)的甲基化和甲基化的保持三个阶段。产生 siRNA 信号的途径很多,包括 RNA 病毒、转座子和重复序列来源或人工导入的 siRNA。在胚乳蛋白 *FWA* 基因的串联重复子中,从头合成 DNA 甲基化需要典型的 siRNA 产生途径,包括核蛋白 AGO4、RDR2 和 SDE4 等蛋白质组分的参与。在由 RNA 病毒产生 siRNA 的途径中,*RDR6*、*SGS3/SDE2* 和 *SDE3* 基因的突变,抑制了 RNAi,同时也抑制了 DNA 的甲基化。*DCL3* 途径与由 24nt siRNA 指导的 DNA 甲基化密切相关,而 AGO1 则与 21nt siRNA 有关,它既能降解 mRNA,也能甲基化病毒或 SPTGS 靶向的 DNA。此外,还有一些产生 siRNA 信号的途径尚待研究。siRNA 产生后,与 DRM1/DRM2 作用,导致 DNA 特异区域的甲基化或组蛋白 H3 亚基第 9 个赖氨酸残基 K(H3K9)的甲基化,但 siRNA 引导 DRM1/DRM2 导致特异区域甲基化的具体机制尚无定论。来自裂殖酵母的实验表明,siRNA 可能与一些蛋白质组分,包括 AGO1、异染色质结合蛋白 Chp1、Tas3 和 DRM 等组成复合体 RITS,在 siRNA 的引导下,复合体使相应的 DNA 区域,包括胞嘧啶、鸟嘌呤核苷酸连续区(称为 CG 岛),CNG(N: A/T/C/G)、CHH(H: A/C/T)中的胞嘧啶核苷 C 发生甲基化或使相应的组蛋白 H3K9 发生甲基化。

在 DNA 水平,异染色质由简并反转录转座子和串联重复序列等组成。异染色质区域的典型标志是由 SU(VAR)39 蛋白介导的组蛋白 H3 亚基的甲基化。一些高度保守的蛋白质与异染色质有关,包括结合于 H3K9 甲基化的异染色蛋白

HP1(heterochromatin protein 1)和负责染色单体的黏连蛋白 cohesin。裂殖酵母中,Swi6 和 Clr 分别是 HP1 和 SU(VAR)3~9 的同源蛋白质。在裂殖酵母中,异染色质沉默基因通常发生在着丝粒、端粒和配型区(mating-type loci)。RNAi 途径的一些蛋白质,如 AGO、Dicer、RdRP 的基因都是单一拷贝基因。在这三种酶突变的 RNAi 突变体中,H3K9 的甲基化程度下降,重复序列相关的 Swi6 蛋白表达量也下降,同时着丝粒外部转座子重复序列表达阻滞程度降低,表明 RNAi 的突变使异染色质阻滞基因表达的程度降低,从而使原本被沉默的基因得到一定程度的表达。RNAi 的突变导致黏连蛋白 cohesin 的表达受阻,使染色体在细胞分裂后期出现拖尾现象,这与异染色质高度保守的蛋白 Clr4 和 Swi6 突变引起的现象一样,揭示了 RNAi 在异染色质形成过程中的作用。RNAi 参与裂殖酵母染色质形成的另一个有力证据来自于对转基因的研究,位于着丝粒重复序列内的转基因通常不表达,但在 RNAi 突变株中,它具有表达活性。同样的现象也发生在配型区,说明 RNAi 在异染色质形成中起着重要作用,RNAi 突变会激活原异染色质区域中沉默基因的表达。对裂殖酵母的生物化学实验分析表明,着丝粒重复序列在 RdRP 的作用下扩增出大量的 dsRNA,dsRNA 被 Dicer 酶加工为 siRNA,siRNA 与一些蛋白质,包括 AGO1、Chp1 和 Tas3 等组成一个核效应复合体 RITS(RNA induced initiation of transcriptional gene silencing),RITS 结合于染色质特定的沉默效应区,其中的 siRNA 引导组蛋白甲基化转移酶对 H3K9 的甲基化。甲基化的组蛋白 H3 亚基被 Swi6 结合而保持其甲基化水平,从而抑制基因的转录活性。

综上所述,siRNA 引发的 DNA 甲基化和异染色质的形成是表观调控的两种重要方式,它们在基因组水平调控基因的表达和调节发育时相,并与肿瘤的发生密切相关。它们的重要区别是后者能够使修饰从起始区域(如启动子区)迅速延伸到其他区域,而 siRNA 引发的 DNA 甲基化仅仅发生在与 RNA 精确互补的 DNA 区域。

三、miRNA 参与基因的表现修饰调节

基因调控过程中的表观修饰调节包括 DNA 的修饰和组蛋白的各种修饰等。在 DNA 的甲基化或去甲基化过程中,受到甲基转移酶(DNMT)作用,能将甲基从一种化合物转移给另一种化合物,其受体包括 DNA、RNA 和蛋白质。因此,甲基化酶在 DNA 和组蛋白的甲基化修饰中发挥重要作用。组蛋白的乙酰化修饰是组蛋白修饰的重要方式,通常情况下组蛋白的乙酰化有利于 DNA 与组蛋白八聚体的解离,核小体结构松弛,从而使各种转录因子和协同转录因子能与 DNA 结合位点特异性结合,激活基因的转录。在细胞核内,组蛋白乙酰化过程与组蛋白去乙酰化过程处于动态平衡,并由组蛋白乙酰化转移酶(histone acetyltransferase, HAT)和组蛋白去乙酰化酶(histone deacetylase, HDAC)共同调控。HDAC 是一类蛋白