

miRNA与LncRNA 相互调控新进展

主 编 姜怡邓 杨晓玲



科学出版社

感谢国家自然科学基金(81560084, 81570452, 81360027, 81360053, 81260105)、宁夏自然科学基金(NZ15209)和宁夏科技领军创新人才项目的资助(2015, 2016)

miRNA 与 LncRNA 相互调控新进展

主 编 姜怡邓 杨晓玲

副主编 张慧萍 赵 巍 马胜超 杨安宁

编 委 (以姓氏笔画为序)

丁 宁	马燕琼	王 楠	毛彩艳
朱光荣	孙 岳	李淑强	杨松昊
张 辉	和杨杨	赵滑国	侯苗苗
徐灵博	高婷婷	郭 伟	谢 琳

科学出版社

北 京

内 容 简 介

疾病的发生及发展是细胞内多个分子、多环节作用失调的结果,蛋白编码基因仅仅是疾病发生过程中复杂的分子网络相互作用的一部分,非编码基因在其中也发挥着重要的调控作用。miRNA 是一类长度 20~24nt 非编码调控单链小分子 RNA,通过靶向结合基因的 3'-UTR 区使基因表达沉默;LncRNA 是一类长度大于 200bp 的长链非编码 RNA,可以通过与互补的 mRNA 形成双链 RNA,影响 mRNA 剪接、加工、转运、翻译与降解等过程,这些 RNA 在机体的生长发育、细胞凋亡、增殖和分化等过程中发挥重要作用。本书就 miRNA、LncRNA 的作用、机制、生物学功能、与疾病的关系及二者间相互调控以及研究方法进行了详细的阐述。

本书可作为 miRNA 与 LncRNA 的病理学与病理生理学专业研究生、科研工作者的参考用书。

图书在版编目(CIP)数据

miRNA 与 LncRNA 相互调控新进展 / 姜怡邓, 杨晓玲主编. —北京: 科学出版社, 2016.6

ISBN 978-7-03-049224-1

I. ①m… II. ①姜… ②杨… III. ①基因表达调控-研究 IV. ①Q786

中国版本图书馆 CIP 数据核字 (2016) 第 147107 号

责任编辑: 王 颖 / 责任校对: 郭瑞芝

责任印制: 徐晓晨 / 封面设计: 陈 敬

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京中石油彩色印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 6 月第 一 版 开本: 787×1092 1/16

2016 年 6 月第一次印刷 印张: 18

字数: 432 000

定价: 99.00 元

(如有印装质量问题, 我社负责调换)

目 录

第一篇 非编码 RNA 的概述

第一章 长链非编码 RNA	1	第三章 长链非编码 RNA 血管相关性	
第一节 长链非编码 RNA 的概述	1	疾病	27
第二节 长链非编码 RNA 的调控机制	3	第一节 血管性疾病危险因素与病理	
第二章 长链非编码 RNA 与表观遗传学		机制	27
调控	12	第二节 长链非编码 RNA 与高脂血症	35
第一节 表观遗传学调控概述	12	第三节 长链非编码 RNA 与肾损伤	40
第二节 长链非编码 RNA 与 DNA 甲		第四节 长链非编码 RNA 与高血压发展	
基化	18	关系	47
第三节 长链非编码 RNA 与组蛋白修饰		第五节 长链非编码 RNA 调控与糖尿病	
(甲基化、乙酰化)	19	血管病变	51
第四节 长链非编码 RNA 与染色质重塑	20	第六节 长链非编码 RNA 与心脏、血管	
第五节 长链非编码 RNA 表观遗传学		发育	60
调控与基因组印记	21	第七节 长链非编码 RNA 与动脉粥样	
第六节 长链非编码 RNA 与其他调控	24	硬化	65
第七节 长链非编码 RNA 的病理生理		第八节 长链非编码 RNA 与心力衰竭	70
意义	24	第九节 长链非编码 RNA 与自噬概述	75

第二篇 小分子 RNA

第四章 miRNA 的概述	81	心脏病	122
第一节 miRNA 概述	81	第二节 miRNA 与心肌纤维化	125
第二节 miRNA 调控方式	87	第三节 miRNA 与心力衰竭	127
第三节 miRNA 与心血管疾病概述	88	第四节 miRNA 与高血压	130
第四节 miRNA 与其他疾病概述	93	第五节 miRNA 与心律失常	132
第五章 miRNA 与动脉粥样硬化	108	第六节 miRNA 与先天性心脏病	134
第一节 miRNA 与动脉粥样硬化发病		第七节 miRNA 与心肌肥厚	135
机制概述	108	第八节 miRNA 和脑卒中	138
第二节 miRNA 与血管相关细胞	109	第九节 miRNA 与血管炎关系	141
第三节 miRNA 与免疫和炎症反应	115	第十节 miRNA 与心肌梗死	143
第四节 miRNA 与脂质代谢	118	第十一节 miRNA 与妊娠高血压疾病	145
第五节 miRNA 在血管疾病中的检测		第七章 miRNA 的调控机制	153
价值	120	第一节 miRNA 与表观遗传学概要	153
第六章 miRNA 与心血管疾病	122	第二节 miRNA 与 DNA 甲基化调控	161
第一节 miRNA 与冠状动脉粥样硬化性		第三节 miRNA 与组蛋白修饰	172

第四节	miRNA 与基因印迹、染色质 重塑	178	第三节	miRNA 与转录因子的调控 机制	192
第八章	miRNA 与转录因子	184	第四节	miRNA 与转录因子结合位点 的预测及调控分析	196
第一节	miRNA 与基因表达调控	184	第五节	展望	206
第二节	转录因子概述及基因表达 调控	186			

第三篇 长链非编码 RNA 与 miRNA 相互调控

第九章	长链非编码 RNA 与 miRNA 相互 作用	209	第四节	长链非编码 RNA 经表观遗传学 机制调控 miRNA	227
第一节	长链非编码 RNA 和 miRNA 简介	209	第十章	长链非编码 RNA 与 miRNA 相互 调控与疾病发展	232
第二节	长链非编码 RNA 与 miRNA 相互关系	213	第一节	长链非编码 RNA 与 miRNA 相互调控与心血管疾病	232
第三节	长链非编码 RNA 与 miRNA 相互调控的机制	215	第二节	长链非编码 RNA 与 miRNA 相互调控与肿瘤	234

第四篇 非编码 RNA 的检测技术

第十一章	长链非编码 RNA 检测技术	237		(cHART)	258
第一节	长链非编码 RNA 的生物信息 学分析及数据库	237	第十节	全局持续测序 (GRO-seq)	259
第二节	长链非编码 RNA 芯片、 RNA-seq 技术	238	第十二章	miRNA 的检测技术与研究方法	261
第三节	长链非编码 RNA 的 RNA 结构 平行分析技术 (PARS)	242	第一节	miRNA 的生物信息学分析及 数据库构建	261
第四节	长链非编码 RNA 结合蛋白免疫 沉淀技术 (RIP)	247	第二节	基因芯片技术及分析	262
第五节	非编码 RNA 沉默和定位分析 (c-KLAN)	249	第三节	Northern 印迹技术	263
第六节	环状染色质构象捕获 (4C)	250	第四节	化学发光标记检测	266
第七节	三维 DNA 的筛选和交联 (3D-DSL)	253	第五节	纳米粒子的荧光共振能量 转移法	267
第八节	RNA 反义纯化 (RAP)	255	第六节	电化学检测	268
第九节	捕获杂交分析 RNA 靶点 (cHART)	258	第七节	原位杂交技术	272
			第八节	实时反转录聚合链式反应	275
			第九节	滚环扩增阳离子共轭聚合物均 相检测 miRNA	277

参考文献	280
------------	-----

第一篇 非编码 RNA 的概述

第一章 长链非编码 RNA

第一节 长链非编码 RNA 的概述

一、LncRNA 的来源

传统的遗传学中心法则报道,遗传信息储存在 DNA 中,蛋白质行使遗传信息的生物学功能,而 RNA 介于 DNA 和蛋白质之间,扮演了桥梁的角色。众所周知,整个基因组中有高达 98% 的 DNA 序列被转录,然而,起到编码蛋白质作用的基因只占不到 2%;其余的超过 90% 的基因组序列被转录成非编码 RNA (non-coding RNA, ncRNA)。最初人们认为,这些 ncRNA 是体内的“转录噪声”,但随着近年来的研究发现,它们参与了体内绝大部分的生物学过程。

一般根据序列的长度不同,可以将 ncRNA 分为两大类:短链非编码 RNA (small ncRNA) 和长链非编码 RNA (long ncRNA, LncRNA)。近 10 年来,piwi 相互作用 RNA (piwi-interaction RNA, piRNA)、小干扰 RNA (small interference RNA, siRNA) 和微小 RNA (microRNA, miRNA) 等短链非编码 RNA 得到了极大的关注和研究。虽然 LncRNA 的研究起步较晚,但它们也在近期越来越受到研究者的青睐。

通过转录真核生物的基因形成蛋白质编码 RNA (protein-coding RNA, mRNA) 与非编码 RNA (ncRNA)。以哺乳动物为例,预计大约有超过 50% 的基因能被表达,但目前为止只发现了约 2% 的基因组负责编码蛋白质,其余的表达基因很可能转录成为 ncRNA,其中包括 siRNA、snoRNA、miRNA 和 piRNA 等在内的短链 ncRNA (small ncRNA) 已经得到广泛研究。尽管如此,对另一类被称为长链非编码 RNA (长度大于 200nt) 的 RNA 分子的研究才刚刚进入发展阶段。研究表明长链非编码 RNA 分子在生命过程中的转录、转录后调控等各个阶段都发挥着重要的生物功能。

Cesana 等发现鼠的长链非编码 RNA 分子 Linc-MD1 作为竞争性的内源 RNA 通过影响 miRNA 与靶基因结合而改变肌肉的分化过程,在后转录水平发挥其调控功能;Huarte 等通过实验观察到 LincRNA-p21 在 p53 抑制基因表达的调控通路中通过与 HnRNP-K 相互作用发挥重要的调控功能。

长链非编码 RNA 分子还可以通过顺式或反式调控影响基因的表达,其在与表观遗传有关的生命过程中同样也发挥了重要作用,如著名的长链非编码 RNA 分子 Xist 及 H19 被发现与 X 染色体失活相关,Jpx 同样也被发现在 X 染色体失活过程中发挥了重要作用,而 Kcnqlot1 被报道可以导致表观遗传基因的沉默,RNCR2 被报道与鼠的视觉细胞分化相关。

研究人员发现长链非编码 RNA 与疾病的发生发展之间存在着密切关系,如长链非编码 RNA 分子 BACE1-AS 被发现与阿尔茨海默病发生发展相关,Gibb 等通过 SAGE 技术对癌症组织和正常组织的对照分析也发现了大量可能与癌症发生相关的长链非编码 RNA,HOTAIR 被发现可以通过染色体重塑引起癌症的转移,因此长链非编码 RNA 分子作为可能的药物靶标被进一步深入研究。

真核基因组不仅是我们之前所认为的简单的、高度有序的基因转录组。而 RNA 不再是仅作用于 DNA 和蛋白质之间的一种信使。伴随着高通量测序的深入应用,真核生物中大量非编码 RNA (ncRNA) 的转录本被发现。研究证实,ncRNA 表现出对蛋白编码基因的复杂调控作用,并参与肿瘤发生发展的一系列病理过程。ncRNA 从长度上分为短链 ncRNA (miRNA、siRNA、piRNA) 和长链 ncRNA (LncRNA)。LncRNA 是指长度大于 200nt 且不编码蛋白质的类似 mRNA 结构的 RNA。近来发现的 HOTAIR 基因是 LncRNA 中的一个重要分子,它位于人类 12 号染色体 HOXC11、HOXC12 基因之间,通过募集组蛋白甲基转移酶 PRC2 复合物作用于 2 号染色体 HOXD 基因簇,并致其甲基化从而反式作用沉默了 HOXD 基因。组织中高表达 HOTAIR 基因往往预示着肿瘤的发生,在结直肠癌、乳腺癌、胰腺癌和肝癌等均有高表达。由 XIST 基因编码产生的 LncRNA Xist 对哺乳动物 X 染色体失活起了关键作用。据报道称,人类常染色体的整个基因组的核苷酸至少在一种细胞中全部被转录。然而人类基因组中大概只有 5%~10% 的基因组可以在转录本的信使 RNA (mRNA) 和 ncRNA 中稳定存在。对于这些转录出来的转录本而言,大多数是不编码蛋白质的。只有约 1% 的人类基因组序列编码蛋白质,余下的 4%~9% 人类基因组序列编码蛋白质转录出来而没有翻译,只有少数的 ncRNA (运输 RNA、核糖体 RNA、LncRNA 和少量的短链 RNA 等) 功能被证实。

大多数的 LncRNA 被认为是处于低进化压力的,只有少数的 LncRNA 展现出种间保守。只有有限的一部分 LncRNA 能解释为在一些种系中快速出现与消失。由于蛋白质编码基因不同,ncRNA 基因不形成大的同源家族。蛋白质编码基因与非编码基因具有完全不同的进化过程,是通过整个或部分基因的复制及后来进化分歧而产生的。

高通量基因测序技术的快速发展使大量的 LncRNA 被发现。目前认为 LncRNA 主要来源于以下途径。①蛋白质基因在多种因素作用下断裂以后形成;②染色质重排过程中 2 个分开的区域紧靠到一起而形成;③短链 ncRNA 中的某段序列多次复制而形成;④非编码基因转录而形成;⑤转录因子中插入一段序列而形成。因此,LncRNA 的发现对生命起源学说提出了挑战,它们具有短侧翼序列,这有利于其维持结构和稳定功能。虽然 LncRNA 保守性很低,但其具有很高的丰度,在编码和非编码序列之间散状或重叠分布,这个复杂体使我们对基因结构的理解从线性水平上升到了分子水平,一个序列被转录成为有义链或无义链、编码或非编码链都是有可能的。许多 LncRNA 只有在特定的发育过程中表达,如在小鼠体内,它们在胚胎干细胞分化过程中特异表达;而在大脑中,它们具有亚细胞精确的定位功能。与编码蛋白质的基因相比较,LncRNA 有不同的保守类型,这有利于保护开放阅读框和功能的保守性,此外还具有不同的表达模式,对大量组织与细胞都有调节作用,这就更增加了研究其复杂生物学功能的迫切性。

二、LncRNA 的分类

尽管在过去 20 多年中,也有 LncRNA 的一些相关报道,但随着高通量测序技术的快速发展才使得从基因水平研究 LncRNA 成为可能。为了更深入地展开 LncRNA 的相关研究,了解其功能及其作用机制,科学家们根据 LncRNA 在基因组中所处的背景及位置,将 LncRNA 分为①基因间 LncRNA (intergenic LncRNA);②内含子 LncRNA (intronic LncRNA);③正义 LncRNA (sense LncRNA);④反义 LncRNA (antisense LncRNA) 4 种类型。

然而,根据它们自身在基因组上相对于蛋白编码基因的位置,也可以将其分为 5 种类型。①正义 LncRNA: 与同一链上另一转录物的 1 个或多个外显子相重叠;②反义 LncRNA: 与另一条链上的转录物外显子相重叠时;③双向 LncRNA: 这类 LncRNA 的表达起始位点与其互补

链上相邻编码转录物的表达起始位点十分靠近；④内含子 LncRNA：来源于编码基因剪接转录物的内含子；⑤基因间 LncRNA：位于两个编码基因的间隔中。反义 LncRNA 因其结构特殊而备受研究人员的关注，它们通过与其他 RNA 形成了 RNA 互补双链、改变染色质重塑等形式调控着靶基因的表达，并且有可能在疾病发生中发挥独特的作用。在转录过程中，除了需要 RNA 聚合酶以外，还需要转录因子和顺式作用元件。

转录因子是一类可以与 DNA 序列发生特异结合的蛋白质，它们具有效应结构域和 DNA 结合结构域等。与蛋白质编码基因相似，反义 LncRNA 基因的启动子也存在一些转录因子的结合位点。Gagnon 等先将成纤维细胞生长因子 2 (fibroblast growth factor-2, FGF-2) 基因启动子区域的反义链与荧光素酶基因共同转染到白血病 T 细胞 (Jurkat T) 中，然后通过启动子区域序列特异突变实验和体外结合实验，证明了 Ets 样的转录因子对反义 FGF-2 具有调控作用。Cawley 等利用高密度寡聚核苷酸芯片分析人的 21、22 号染色体上 p53、SP1 和 c-Myc 的转录因子。结果发现，一些 ncRNA 的基因和蛋白质编码基因的启动子区域受到共同转录因子的调控。这些现象说明，有些反义 LncRNA 基因可能通过结合与正义链相同的转录因子从而调控正义链的转录。

第二节 长链非编码 RNA 的调控机制

一、LncRNA 转录水平的调控

无论是原核生物还是真核生物，转录水平的调节都是基因表达调控最重要的环节，LncRNA 可以通过多种方式来调节基因的转录，如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。

近年来，在对大量基因组序列研究的过程中，研究人员得出一个结论：在编码蛋白质的基因以外，有很多的基因组都被转录了。通过对染色质的分析发现，哺乳动物体内 LncRNA 可达 1000 多种，都是通过蛋白质编码序列转变为非编码序列后转录而得到，或由去除了外显子的基因序列转录而来，其表达具有组织和时空特异性。与编码蛋白质的基因相比，LncRNA 序列的保守性较差。与大多数非编码 RNA 一样，LncRNA 仅在 RNA 的二级结构和启动子区有进化保守性，多数由 RNA 聚合酶 II 转录生成：①LncRNA 可以通过诱导组蛋白修饰使染色质发生重构、干扰转录因子与启动子结合等方式直接抑制邻近蛋白质编码基因表达，并且 LncRNA 转录停止后，其抑制作用依然存在；②LncRNA 可通过与蛋白结合伴侣作用或调节蛋白的亚细胞定位，影响蛋白质的活性，RNA 结合蛋白 TLS 与从细胞周期蛋白 D1 (CCND1) 启动子转录来的 LncRNA 结合后变为有活性构象，抑制 CREB 结合蛋白 (CBP) 及组蛋白乙酰基转移酶 p300 的活性，从而阻断 CCND1 的转录；③LncRNA 作为重要的调节因子，与靶蛋白结合后影响转录，LncRNA 可使聚梳 (PcG) 蛋白和胸板 (TrxG) 蛋白聚集到目的基因上，抑制或激活目的基因的表达，同时有研究显示，LncRNA AIR 可募集组蛋白甲基转移酶 G9a，使启动子区组蛋白甲基化，引起转录沉默；④LncRNA 结合转录抑制复合物后有抑制肿瘤形成的作用。

LncRNA-p21 与抑制复合物不均一核糖核蛋白 K (hnRNP-K) 的准确结合及正确定位可使细胞发生凋亡，这是最新发现的 p53 基因发挥抑癌作用的重要机制。此外，LncRNA 还可影响 RNA 的加工处理过程，并且是非编码 RNA 的前体、细胞骨架及有丝分裂纺锤体的结构成分。

(一) 转录干扰

转录干扰在真核生物与原核生物中均存在，应用原子力显微镜在大肠杆菌中观测到了这

种碰撞。天然反义转录物 (natural antisense transcript, NAT) 是指天然情况下生物体内生成的反义 RNA, 它们可以与互补 RNA 通过碱基配对, 形成双链 RNA (double strand RNA, dsRNA)。转录干扰中 RNA 聚合酶碰撞的概率往往与 NAT 重叠区的长度和方式有关, 随着重叠区增长碰撞概率增大。5'端是转录的起始点, 在“头对头”NAT 中, 转录干扰将会导致转录起始的抑制。如果一个链率先转录, RNA 聚合酶会覆盖另外一条链的启动子, 或改变另外一条链启动子区的拓扑异构结构, 从而影响另外一条链上转录起始复合物的装配; 如果两条链都已经开始转录, 常常会导致双方转录终止。然而完全重叠的 NAT, 即使重叠的区域较小, 也会影响较大。转录干扰可以产生长度不一的 mRNA, 这也许是 Northern 印迹杂交中会出现异常带的一个原因。

LncRNA 结合转录因子, 可以直接顺式调节基因的转录, 而不是通过与 DNA 结合而发挥转录调控作用。例如, 电离辐射诱导的细胞周期蛋白 D1 (cyclin D1 CCND1) 启动子上游转录出的 LncRNA 与脂肪肉瘤易位蛋白 (translocation liposarcoma protein, TLS) 结合, 从而产生了变构效果, 结合并抑制组蛋白乙酰转移酶, 致使 CCND1 转录下调 (图 1-1)。

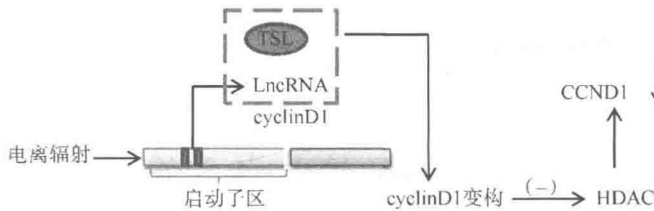


图 1-1 电离辐射影响细胞周期蛋白 D1 转录降低的机制

另外一些基因的 mRNA 的内源性互补转录本——NAT 或 AS LncRNA) 在转录时, 与正义转录本 (mRNA) 的重叠区较大, 两个 RNA 聚合酶相互碰撞导致转录起始点受阻, 产生一条链转录而另一条链转录受到抑制的转录干扰现象, 尤其见于“尾对尾”转录的反义转录本中。

1. 通过转录碰撞实现转录抑制 在转录延伸过程中, 催化正、反链合成的 RNA 聚合酶复合物在互补区域发生相互碰撞, 导致转录阻滞的现象称为转录碰撞 (transcriptional collision)。在 2002 年, Prescott 等在出芽酵母中发现了这一现象。随后, Crampton 等利用原子力显微技术在大肠杆菌中观察到了这一现象。Hobson 等进一步研究发现, 当碰撞发生在体外时, RNA 聚合酶 II 停止延伸, 但并不发生解离; 而当碰撞发生在体内的时候, 由于在细胞内缺乏有效的 RNA 聚合酶 II 多聚泛素化作用, 这使得 RNA 聚合酶 II 的半衰期延长, 增加了碰撞的可能性。生物信息学研究发现, 在人和小鼠中, 正、反转录产物的重叠区域的长度与反义转录物的表达水平呈负相关, 这说明发生转录碰撞的可能性与重叠区域的长度有关。

2. 竞争性结合转录因子实现转录抑制 LncRNA 调节转录因子的结合与装配 在转录过程中, 除了需要 RNA 聚合酶以外还需要转录因子和顺式作用元件。

转录因子是一类能够与 DNA 序列发生特异性结合的蛋白质, 它们具有 DNA 结合结构域和效应结构域等。与蛋白质编码基因相似, 反义 LncRNA 基因的启动子也有一些转录因子的结合位点。Gagnon 等先将纤维母细胞生长因子 2 (fibroblast growth factor-2, FGF-2) 基因启动子区域的反义链与荧光素酶基因共同转染到白血病 T 细胞 (Jurkat T) 中, 然后通过启动子区域序列特异突变实验和体外结合实验, 证明了 Ets 样的转录因子对反义 FGF-2 具有调控作用。Cawley 等利用高密度寡聚核苷酸芯片分析人的 21、22 号染色体上 p53、SP1 和 c-Myc 的转录因子。结果发现如下, 一些 ncRNA 的基因和蛋白质编码基因的启动子区域受到共同的转录因子的调控。这些现象说明了, 有些反义 LncRNA 基因可能通过结合与正义链相同的转录因子从而调控正义链的转录。Evf-2 属于超保守的 LncRNA 中的一员, 是同源盒转录因子 Dlx2 的辅因

子,由 Dlx-6 和 Dlx-5 基因远端的增强子元件转录。Evf-2 可以招募 Dlx2 到增强子,激活 Dlx-6 和 Dlx-5 基因的表达,从而调控前脑的发育和神经发生。在 DNA 受损伤时,周期蛋白 D1 基因的启动子可以转录一种低拷贝的 LncRNA,后者可以结合到 CCND1 基因的 5'端调控区,选择性招募 DNA 损伤信号转录调节传感器蛋白 TLS 到 CCND1 的启动子位点,调控 CCND1 基因的表达,并抑制 cAMP 应答元件结合蛋白 CREB 和组蛋白乙酰转移酶 p300 的活性。哺乳动物中, PcG 蛋白质可以结合并沉默上千基因,而 TrxG 蛋白质则能够维持基因的活化状态,它们与目标位点的结合也均是由 LncRNA 指导。而在酵母 IMD2 和 URA2 基因位点,蛋白质编码基因和 LncRNA 基因具有相同的启动子,上游 LncRNA 基因可以通过竞争转录因子,抑制下游蛋白质基因的表达。

人类二氢叶酸还原酶基因 DHFR 上游的一个小启动子,能够转录 LncRNA,后者则能够通过启动子 DNA 形成三链复合物,从而阻止转录因子 TF II D 与启动子的结合,抑制 DHFR 基因转录。因为真核生物染色体中存在许多类似的三链结构,这种启动子的调控方式可能在真核生物中非常普遍。线状染色质的端粒是 DNA-蛋白质复合体,端粒区可以转录 LncRNA-TERRA (telomeric repeat containing RNA, TERRA),由此推测酵母端粒 LncRNA 也能够和端粒 DNA 形成 RNA-DNA 杂交分子,从而抑制端粒酶活性并导致端粒缩短。

3. 通过竞争性顺式作用元件实现转录抑制 顺式作用元件是可以与特异转录因子结合并影响转录活性的 DNA 序列。起到正性调控作用的顺式作用元件包括增强子和启动子,而起负性调控的元件则是沉默子。转录因子 PU.1 是正常造血过程中很重要的调控因子,它的表达水平常常与细胞命运密切相关,其功能一旦发生紊乱,即便是数量的稍微减少,也会导致白血病和淋巴瘤的发生。在人和小鼠中, PU.1 基因的表达受到-17kb 到-14kb 的上游调控元件 (upstream regulatory element, URE) 的调控。通过同源重组去除 URE 以后,野生型小鼠 PU.1 的表达水平将降低 80%,同时促进了白血病和淋巴瘤的发生。在 PU.1 基因位点上,除了编码其本身的 mRNA 外,还可编码来源于同一内含子、启动子的反义 LncRNA。经过研究证实,正反转录物的表达依赖于 URE, URE 与这些转录物基因的启动子可以相互作用。当用 siRNA 沉默反义 LncRNA 表达时, PU.1 的表达水平将会增加,这说明反义 LncRNA 会与 PU.1 竞争 URE,从而负性调控 PU.1 的表达。这就从另一个方面提供了一种可能治疗白血病和淋巴瘤的方法。

4. LncRNA 调控 RNA 聚合酶 II 人类短散布核元件序列,如 Alu,几乎占到了基因组的 10%。这些元件可以在遇到热刺激等外界胁迫时由 RNA 聚合酶 III 转录 LncRNA,后者含有类似蛋白质转录因子的两个松散结构域,能够和 RNA 聚合酶 II 结合,抑制前转录活性复合体的形成,阻止转录的起始,并迅速改变基因表达模式。这种功能性重复结构域在其他一些 LncRNA 也被发现,推测这种机制可能在生物中具有一定的普遍性。由于很多 RNA 聚合酶 III 转录的 ncRNA 可以起到调节 RNA 聚合酶 II 活性的作用,所以有人提出了“cogene/gene”调节网络的概念。

(二) 与 DNA 形成异源多聚体

LncRNA 与不同功能的几类效应分子:抑制性复合物蛋白质如聚梳蛋白基团、激活性复合体如三胸苷蛋白等构成大分子复合体,然后与靶基因相结合,形成了蛋白质-RNA-DNA 异源多聚体调节基因的转录。在 DNA 损伤的情况下位于 CDKN1A 基因上游的 LncRNA-p21 被诱导转录,其作为经典 p53 通路的转录抑制因子,通过结合和修饰核内非均一核糖核蛋白 K (hnRNP-K),从而作用并抑制 p53 基因,引发细胞凋亡。有研究人员认为: TUG1 和 NEAT2 激活或抑制基因的表达也是通过与 DNA 形成异源多聚体实现的。

二、LncRNA 转录后水平调控

LncRNA 在转录后水平可以通过与互补的 mRNA 形成双链 RNA, 影响 mRNA 剪接、加工、转运、翻译与降解等过程, 从而起到调节基因表达的作用。人急性粒细胞白血病细胞 HL-60 细胞系中 PU.1 的 mRNA 水平很高, 而蛋白质水平却不高, 是由于 PU.1 的 asRNA 与其 mRNA 结合使翻译受到抑制。LncRNA 还可以互补结合 mRNA, 掩蔽其 3'UTR 部位的靶 miRNA 结合位点, 从而有助于该 mRNA 的稳定并促进其翻译, interferon- α 1 的 asRNA 即通过掩蔽其 mRNA 的 miR-1270 结合位点增强其 mRNA 的稳定性。

(一) mRNA 降解

mRNA 的丰度与蛋白质的产量有直接的关系, 而影响其丰度的主要原因有降解速率和转录量。mRNA 的转录量主要通过转录水平调控及转录后加工决定, 而 mRNA 通过多种途径被降解并且影响其丰度。无义介导的 mRNA 降解 (nonsense-mediated mRNA decay, NMD) 是使 mRNA 降解的重要途径之一。RNA 结合蛋白 Stau1 (Staufen1) 则通过直接结合一种称为移码增加蛋白 1 (up-frameshift protein 1, Upf1) 的 NMD 因子, 并将其募集到 mRNA 的 3'-非翻译区 (untranslated region, UTR), 引起 mRNA 的降解。这种机制被命名为由 Stau1 介导的 mRNA 降解 (stau1-mediated mRNA decay, SMD)。Stau1 通过 SMD 途径调控着大量转录物的水平。Gong 等发现, 一种称为半 Stau1 结合位点 (Half-Stau1-binding site) RNA 的 LncRNA, 通过与 mRNA 的 3'UTR 的 Alu 元件不完全配对, 形成 Stau1 结合位点, 促进了 Stau1 与 mRNA 结合, 导致 mRNA 降解。这一发现表明 LncRNA 在调控转录物丰度方面也发挥着作用。

(二) 翻译调控

LncRNA 也在翻译水平发挥表达调控作用。 β -分泌酶 1 反义转录物 (BACE1-antisense transcript, BACE1-AS) 是 β -分泌酶 1 (β -site APP cleaving enzyme1, β -secretase 1, BACE1) 基因的天然反义转录物。如果 BACE1-AS 与正义 BACE1 mRNA 相结合将增强 BACE1 mRNA 的稳定性, 从而增加 BACE1 蛋白的表达量, BACE1 蛋白通过切割 β 淀粉样肽前体, 产生更多的 β 淀粉样肽。由于脑内存在很多 β 淀粉样肽斑块是老年痴呆症的一个显著特点, 因此, Faghihi 等认为 BACE1-AS 直接参与老年痴呆症的病理过程。此外, LncRNA-p21 也参加翻译调控, 当细胞中缺少一种被称为人类抗原 R (human antigen R, HuR) 的 RNA 结合蛋白时, LncRNA-p21 在细胞中稳定存在并且不断地积累, 它与靶 mRNA 结合从而抑制后者的翻译。

(三) 剪接调控

LncRNA 不但在转录调控中发挥作用, 还是 mRNA 前体剪接的调控因子。例如果蝇中的热休克 RNA ω -n (heat shock RNA ω -n, hsr ω -n) 和人类中的卫星 III (satellite III, sat III) 重复序列的转录物通过隔离剪接因子调控 mRNA 前体的剪接。另一种被称为肺腺癌转移相关转录物 1 (metastasis-associated in lungadenocarcinoma transcript 1, MALAT1) 的 LncRNA 已被发现在多种癌症中异常表达, 它由其初级转录物的 3'端加工而来, 主要位于剪接斑点 (splicing speckle)。它与丝氨酸/精氨酸 (serine/arginine, SR) 剪接因子相互作用, 并且调控剪接因子在剪接斑点中的分布和磷酸化水平, 继而改变 mRNA 前体的选择性剪接模式。Bernard 等发现, MALAT1 在神经元中是高表达的, 它通过调控 SR 剪接因子影响突触形成相关基因的表达。此外, MALAT1 还与剪接因子反式激活应答 DNA 结合蛋白 43 (transactive response DNA binding

protein-43, TDP-43) 相互作用, 影响 mRNA 前体的选择性剪接。

三、LncRNA 表观遗传学水平调控

(一) 染色质修饰与 DNA 甲基化

染色质的状态是转录活性的关键决定因素之一, 其涉及核小体结构和组蛋白修饰等。最近的研究表明, 一些 LncRNA 通过影响染色质的状态调控了基因的表达。通过募集 PRC2 复合体, 反式抑制 HOXD 位点的转录, 同源异型框基因反义基因间 RNA (HOX antisense intergenic RNA, HOTAIR) 转录自同源异型框基因 C (homeobox C, HOXC) 位点。Tsai 等的研究表明, HOTAIR 作为骨架分子从而发挥作用, 其两端结合不同的组蛋白修饰复合体——5'端结合 PRC2 复合体, 而 3'端结合了 LSD1/CoREST/REST 复合体, 即由阻遏元件 1 沉默转录因子辅阻遏物 (corepressor for repressor element-1 silencing transcription factor, CoREST)、阻遏元件 1 沉默转录因子 (repressor element-1 silencing transcription factor, REST) 和赖氨酸特异性脱甲基酶 1 (lysine-specific demethylase 1, LSD1) 3 种蛋白质结合而形成的复合体。前一个复合体具有促甲基化作用, 后一个复合体具有脱甲基化作用, 由它们分别决定不同靶基因的特异性组蛋白修饰模式 (图 1-2)。HOTAIR 的过表达, 将会引起 PRC2 复合体在全基因组范围内的重新定位, 使若干抑癌基因沉默, 终将促进乳腺癌的转移。而另一种 LncRNA——INK4 位点反义非编码 RNA (antisense non-coding RNA in the INK4 locus, ANRIL) 是周期蛋白依赖性激酶抑制因子 2B (cyclin-dependent kinase inhibitor 2B, CDKN2B, 即 p15 INK4b) 基因的反义转录物, 参与了抑癌基因周期蛋白依赖性激酶抑制因子 2A (cyclin-dependent kinase inhibitor 2A, CDKN2A, 即 p16 INK4a) 基因的沉默。ANRIL 结合并且募集 PRC1 与 PRC2 复合体, 导致染色质状态发生改变, 从而抑制 p16 INK4a 基因的表达。

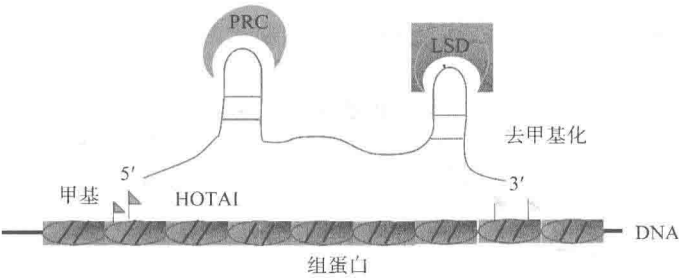


图 1-2 HOTAIR 作为组蛋白修饰复合体的骨架分子调控组蛋白甲基化和去甲基化
HOTAIR 的两端分别结合不同的组蛋白修饰复合体而使组蛋白呈现不同的甲基化状态

(二) X 染色体失活

起始发生在胚胎发育的早期及 X 染色体失活的选择, 这一过程由 X 失活中心 (X inactivation center, Xic) 所控制, 其是一种反义链转录调控模式。Xic 是一个顺式作用位点, 包含辨别 X 染色体数目的信息和 Xist 基因, 前者可保证仅有一条染色体具有活性, 但具体机制尚不明确, 后者缺失将导致 X 染色体失活失败。

X 染色体失活是由标志性的大约为 17kb 大小的 LncRNA——Xist 顺式作用介导, 它在哺乳动物的 X 染色体失活中非常重要。与大多数转录本不同的是, Xist 并不输送至胞质, 其位点上一段 1.6kb 的内部非编码转录本——RepA 可以募集 polycomb 抑制性复合物 2 (PRC2), 即染色质重构复合体, 并包裹在合成它的 X 染色体的某个 RNA 结构域上, 在即将失活的染色体表

面形成“外套”（coating）。伴随着 Xist 在 X 染色体上的扩展，大量的组蛋白甲基化（如 H3K27 三甲基化）即刻发生（这对 X 染色体失活的维持和建立有重要的作用），从而使 X 染色体失活，但由于反义转录本 Tsix 可以限制 Xist 的活性，使剩余的活化 X 染色体还存在有 PRC2。Xist 介导的 X 染色体失活的特征总结来说就是丧失激活性的组蛋白修饰（如乙酰化）而获得抑制性的组蛋白修饰（如 H3K27 的三甲基化），最后，导致失活染色体上的许多 CpG 岛启动子区甲基化，失活的染色体依旧持续合成 Xist，以维持本身的失活状态，但有活性的 X 染色体如何阻止 XistRNA 的具体结合机制还不明确。令人感到有趣的是，即使在细胞分裂中期，该“外套”始终以顺式方式与失活的染色体相连，经多次细胞分裂后，X 染色体失活始终存在，表明该 RNA “外套”与组蛋白修饰及 DNA 甲基化共同构成表观遗传学的记忆，从而使失活的染色体进行有丝分裂。但在体细胞，Xist 基因的缺失并不影响基因沉默，说明一旦建立了记忆，LncRNA 诱导的沉默结构域就不再是必需的了。例如，一转录自 HOXC 基因座的 LncRNA——HOTAIR 具有类似的机制，它能够募集染色质重构复合体 2（PRC2）并将其定位到 HOXD 位点，从而诱导了 HOXD 基因座上长达 40kb 的转录本发生表观遗传沉默（图 1-3）。同样，Kcnq1ot1、Air、这些 LncRNA 都可以通过募集相应的重构复合体，利用甲基转移酶 Ezh2、G9a、Mll 或 PcG 等实现表观遗传沉默。这一机制将染色质重构复合物与发育相关的修饰联系起来。

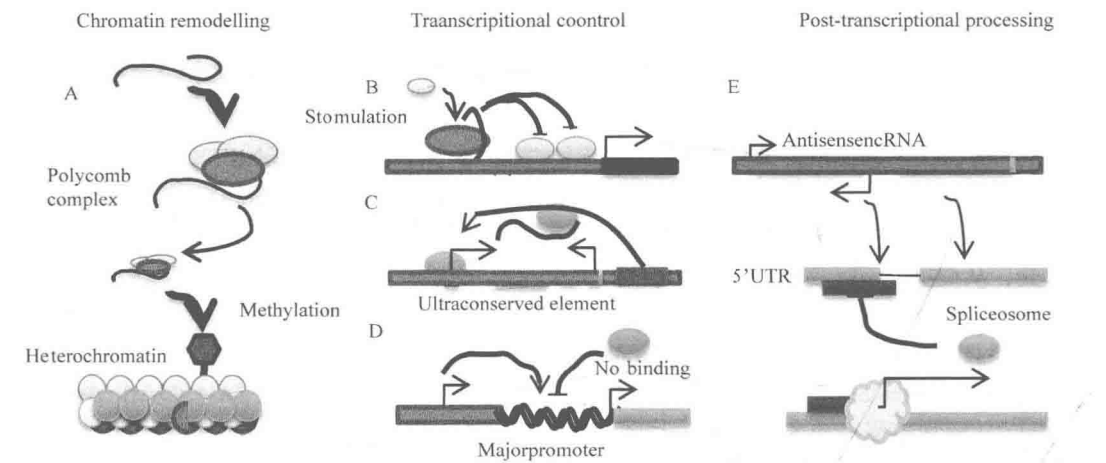


图 1-3 LncRNA 在 3 个层面上实现对基因表达的调控

另有研究人员报道：X 染色体失活依然存在另一机制，即与一种调控性短链 ncRNA 有关。Xist 和反义链转录本 Tsix 退火复性后可形成 RNA 二聚体，在 Dicer 酶作用下，被加工剪切成 24~42nt 大小的调控性 siRNA，这一机制对于失活的 X 染色体上的抑制性异染色质的修饰是非常必需的，因为 X 染色体失活过程中的三甲基色氨酸、组蛋白和赖氨酸等都需要这些 siRNA，并且 X 染色体激活时 Xist 启动子区的 CpG 岛甲基化同样也需要它们。从这一机制我们或许可以推测 LncRNA 和短链 ncRNA 在染色质重塑方面的作用与联系，同时提示我们 RNA 调控存在着一个更为复杂的、交互的网络。

（三）基因组印记

所谓基因组印记，是指二倍体哺乳动物的每一个基因都是双拷贝，一些基因的表达取决于来自父本还是母本，该现象涉及基因表达调控的遗传。基因组印迹是指来自父方与母方的等位基因在通过精子和卵子传递给子代时发生了修饰，使带有亲代印记的等位基因具有不同的表达特性，这种修饰常为 DNA 甲基化修饰，也包括组蛋白乙酰化、甲基化等修饰。

印迹基因在同源染色体上通常以基因簇的形式出现，这些印迹基因簇所对应的染色体区域

被称作印迹区（imprinted region）。印迹基因的表达或沉默是由印迹控制区（imprinting control regions, ICRs, 又称作印迹中心 imprinting centers, ICs）通过顺式作用来调控的, ICRs 的本质是一段差异甲基化区域（differentially methylated regions, DMRs），它的调控作用能够覆盖很长的染色体区域。研究人员发现，每个基因印迹区中至少含有 1 个 ncRNA 基因。

LncRNA 的转录是印迹基因簇调控印记表达的一种重要机制，目前的研究表明，大部分的印迹基因簇的调控都是通过这一机制来完成的。其中代表性的基因包括 BWS 印迹区的 KCNQ1OT1 基因和 AIR-IGF-2R 印迹区的 AIR 基因。以 KCNQ1OT1 为例，其启动子位于 KCNQ1 基因的第 10 内含子内，并且与一个母源高度甲基化的 CpG 岛（ICRs）重叠。因此在 KCNQ1 位点上，这个反义 ncRNA 基因为母源沉默、父源表达。当父源染色体上的这个区域被缺失后，这个印迹基因簇的印迹表达状态便丧失，所有基因恢复双等位基因表达（biallelic expression）。而通过插入转录终止信号将这个 ncRNA 截短后，该印迹基因簇同样失去印迹表达，表明 KCNQ1OT1 的转录产物本身直接参与了印记调控。研究发现，KCNQ1OT1 的作用机制类似于 HOTAIR，能将 Eed-Ezh2 多聚复合物募集到父系染色体 KCNQ1 位点，通过抑制组蛋白甲基化调控基因沉默。

通过研究发现许多印迹基因对胚胎及胎儿出生后的生长发育具有重要的调节作用，对行为和大脑的功能也产生很大的影响。印迹基因的异常表达如印记基因突变、印记丢失等常引发伴有表型缺陷和复杂突变的多种人类疾病，如人类 15 号染色体 q11-q13 印迹区的缺失易引起 2 种迥然不同的神经疾病，Angelman 综合征和 Prader-Willi 综合征。另外，位于 11 号染色体上的 CDKN1C2 和 IGF-2 个印迹基因的错误表达会引起 BWS 综合征等。

H19、X 染色体特异性失活转录物（X inactivation specific transcription, Xist）等多种 LncRNA 也都参与了基因印记。

胰岛素样生长因子 2（insulinlikegrowth factor 2, IGF-2）基因和邻近的 H19 基因均位于 11p15.5，两个基因间相距仅 90kb。H19 在脊椎动物胚胎发育期是高表达的，但出生后大多数组织中的 H19 表达迅速下调。IGF-2 是一种重要的生长因子，在脊椎动物中是高度保守的。在胚胎发育时期，同一发育阶段的相同组织中，这两种基因的表达模式基本相似，它们的表达受到共调控。在大多数的胚胎组织中，IGF-2 和 H19 呈现单等位基因表达（monoallelic expression）。IGF-2 和 H19 基因之间存在着印记控制区（imprinting control region, ICR）。在母源染色体上，ICR 是未被甲基化的，它作为绝缘子（insulator）和转录因子 CTCF（CCCTC-binding factor）相结合，从而阻断下游增强子对 IGF-2 基因的结合，而使增强子作用于 H19 启动子，从而促进了 H19 的表达；然而，在父源染色体上，由于 ICR 是甲基化的，不能和转录因子 CTCF 结合，因此也无法发挥绝缘子的作用。下游增强子只与 IGF-2 基因启动子结合而不与 H19 基因启动子结合，结果促进了 IGF-2 的表达，抑制了 H19 基因的表达（图 1-4）。

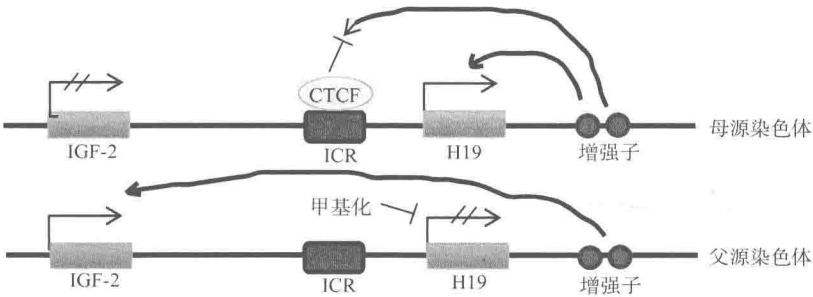


图 1-4 H19 和 IGF-2 基因的印记机制

在母源染色体上 ICR 不被甲基化，所以可以结合 CTCF，结果造成 H19 基因表达，而 IGF-2 基因不表达；在父源染色体上 ICR 被甲基化，结果造成这两种基因的表达情况刚好相反

Xist 是一种被人们所熟悉的 LncRNA，它在 X 染色体失活中具有很重要的作用。在雌性哺乳动物中，X 失活中心（X inactivation center，Xic）控制 2 条 X 染色体中的一条染色体沉默，从而维持剂量补偿效应。Xist 基因的 5'端编码一种称为重复 A（repeat A，RepA）的 LncRNA，它能够与多梳抑制复合体 2（polycomb repressive complex 2，PRC2）结合形成复合体，并移动至 Xic，然后激活 Xist 的表达。随着 Xist 的转录并覆盖 X 染色体，大量的组蛋白被甲基化，最终导致了 X 染色体的失活。另一方面，Xist 的活性受其反义转录物 Tsix（反向书写 Xist）的调控，Tsix 能够有效阻断 Xist 的积累。由此可知，Tsix 的转录对维持 X 染色体的活性是必需的。

(四) 泛素化修饰

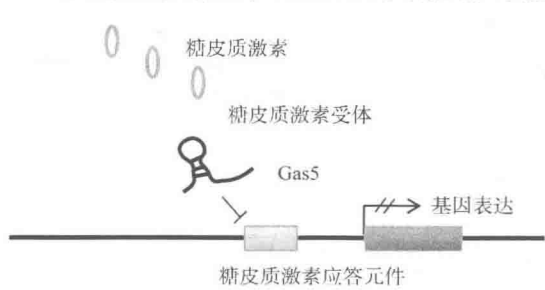
泛素（ubiquitin）是一种由 76 个氨基酸构成的、在真核生物中广泛存在的，并且具有高度保守性的多肽。一个或多个泛素分子在一系列酶的作用下与底物蛋白质分子发生共价结合的翻译后修饰过程称为泛素化修饰（ubiquitination/ubiquitylation）。泛素化修饰最早被用来标记靶蛋白，使之被蛋白酶体识别并降解，整个过程涉及泛素分子、底物蛋白多种酶系统。

(五) LncRNA 去分化和多能性的诱导

与多能性相关的长链基因间的非编码 RNA（long intergenic non-coding RNA，LincRNA）最早发现于鼠胚胎干细胞。Loewer 等研究发现与胚胎干细胞（ESCs）相比较，在诱导多能干细胞（iPSCs）中检测到 10 个的表达上调 LincRNA，推测 iPSCs 的形成是通过这些 LincRNA 与多能性因子相互作用实现的。例如，LincRNA-RoR（receptor super family related to the retinoic acid receptors，RoR）在体细胞去分化并诱导 iPSCs 的过程中高表达，能靶定结合关键的多能性因子 Oct4、Sox2 和 Nanog 的启动子邻近区域（PrPr 区域），从而促进其表达而诱导细胞去分化和多能性的产生。因此，RoR 在体外可以促进多种细胞系去分化产生 iPSCs。

(六) 细胞周期调控

有相关研究报道，LncRNA 具有调控细胞生长的作用，这主要是通过调控细胞周期和细胞凋亡来实现的。生长阻滞特异转录物 5（growth arrest-specific transcript5，Gas5）是哺乳动物细胞凋亡和生长的关键调控因子。通过模拟糖皮质激素应答元件 Gas5 结合糖皮质激素受体（glucocorticoid receptor）的 DNA 结合结构域，阻止了糖皮质激素受体与糖皮质激素应答元件的相互作用，从而抑制了下游基因的表达，促进了细胞凋亡的发生。这就是 LncRNA 作为诱饵分子的一个典型例子（图 1-5）。



糖皮质激素

糖皮质激素受体

Gas5

糖皮质激素应答元件

基因表达

图 1-5 Gas5 作为诱饵分子的机制

Gas5 与糖皮质激素应答元件竞争性结合糖皮质激素受体，从而阻止下游基因的表达

Huarte 等发现，p53 直接诱导长链基因间非编码 RNA p21（long intergenic ncRNA p21，LincRNA-p21）的表达。LincRNA-p21 与核不均一核糖核蛋白 K（heterogeneous nuclear ribonucleoprotein-K，hnRNP-K）相互作用后可抑制 p53 信号通路下游基因的表达，从而调控 p53 介导的细胞凋亡。另一个参与了细胞周期调控的 LncRNA，是 DNA 损伤活化 p21 的相关的非编码 RNA（p21 associated ncRNA DNA damage activated，PANDA）。当 DNA 损伤发生时，p53 结合周期蛋白依赖性激酶抑制因子 1A（cyclin dependent kinase inhibitor 1A，CDKN1A，即 p21）基因位点，然后激活 PANDA 的表达，而 PANDA 可通过结合核转录因子 Yα 亚基（nuclear transcription factor Y subunit α，NF-YA）而

阻止凋亡相关基因的表达,从而延长细胞的生存时间。

四、LncRNA 可能的结构机制与展望

有研究报道称:LncRNA 可能不存在于核糖体样核糖核蛋白复合物(RNPs)中。核糖体是迄今为止已发现的唯一的结构大于 1kb 的 RNA,提示我们,LncRNA 与核糖体在结构组成上是否会有相似之处?最近对类固醇激素受体 RNA 激活剂(SRA) LncRNA 结构研究显示,在整体结构上, RNA 的二级结构与核糖体相似。目前为止,没有关于 SRA LncRNA 或者其他 LncRNA 三级结构的信息,也不确定 LncRNA 是存在于核糖核蛋白复合物中,还是作为独立的 RNA 存在着。

通过与核糖体相比,研究人员发现了一些含有特异性蛋白质的 LncRNA 复合物。在人类全基因组中,蛋白编码基因的总数据估计能达到 21 000。因为多数蛋白质都存在于细胞质中,从而推测胞核中蛋白的数量,即 $n_{\text{protein}}/n_{\text{nucl}}$ 小于 21 000。而已鉴定 LncRNA 多数位于细胞核中,根据保守估计, $n_{\text{LncRNA}}/n_{\text{nucl}}$ 大于 3000,由此 $n_{\text{LncRNA}}/n_{\text{protein}}$, n_{nucl} 小于 1/7。许多 LncRNA 与核糖体大小相当甚至比核糖体还大。以核糖体本身为例,每一个亚基中 RNA 分子与蛋白分子的比例 $n_{\text{RNA}}/n_{\text{Rp}}$ 大约为 1/25。因此,在人类基因组中,即使每一个特异性编码的蛋白都与一个 LncRNA 形成一个复合物,LncRNA 在结构组成上能仍然不可能与核糖体完全相似。迄今为止,没有足够特异的蛋白质能让每一个 LncRNA 都形成核糖体样复合物。由此,存在于胞核中的 LncRNA 不太可能存在于核糖体样 RNP 中。

大量的 LncRNA 也形成与端粒酶 RNaseP、RNA 和 I 类、II 类内含子结构类似的复合物。尽管据推测 LncRNA 复合物的结构与核糖体不同,但却并不排除它与端粒酶 RNA、RNaseP 或 I 类、II 类内含子相似。以“RNaseP 样”复合物为例,LncRNA 是高度结构化和紧密的,含有一个主要的蛋白结合位点,能与各种蛋白结合。经实验证实,SRA 的二级结构是高度组织化的,提示它很可能具有这种结构;而某些 LncRNA 的结构分散,没有一个紧密的核心,它可能含有几个不同的蛋白结合位点,作为具有柔韧性的绳索结构,该结构与端粒酶 RNA 类似;而另一些 LncRNA 也可能是一个独立的、高度结构化的 RNA,与 I 类和 II 类内含子相似。这种 LncRNA 可能只是在需要的时候与蛋白短暂地结合;也可能存在一些高度无序的 LncRNA,含有松散的蛋白结合结构域。

基于 RNA 分子的序列、二级及三级结构,出现了多元化的 RNA 机制及它们的组合机制。在以序列为基础的机制,如 siRNA 介导的 RNA 干扰和 miRNA 介导的 RNA 沉默中, RNA 起到了非常微小的结构作用。它的主要作用是增加这些进程的序列特异性,让 RISC 复合物找到它的作用靶点,进而触发一个主要以蛋白为基础的调节机制。

在过去的 10 年中,发现了一个新的调节机制,这个机制几乎完全是基于 RNA 的二级结构。在核糖开关 RNA 系统中,两个二级结构彼此竞争从而控制了转录的终止,一个代谢产物的存在与否选择这两种结构中的一种,决定了开关基因表达的启闭。例如,SAM-I 核糖开关中,代谢产物(SAM)的存在能引起 RNA 折叠成一个紧密的配体,有利于转录终止子螺旋结构的形成,从而关闭 SAM 合成酶的基因表达。在缺乏该代谢产物时,形成另一种螺旋结构,抑制转录终止以 RNA 三级结构为基础的往往是变构机制,可被称为“构象选择”或“诱导契合”。在诱导契合中,一个事件,如蛋白或配体结合,可以引起一个大的构象变化。配体或蛋白的结合将打破两种构象形成的动态平衡,而趋向于形成其中一种。以核糖体为例,不同构象的涨落常常同时存在于不同的时相。GTP 水解或蛋白结合或使这种波动同步,推动着不同构象间动态平衡和能量变化相适应,使核糖体在延伸周期中顺利前进。子螺旋结构的形成,开启了基因表达。

第二章 长链非编码 RNA 与表观遗传学调控

第一节 表观遗传学调控概述

所谓表观遗传学，就是不改变基因的序列，通过对基因的修饰来调控基因的表达。所以，基因表达的表观遗传学调控，就是通过各种表观遗传的修饰方式来对基因进行调控。目前，已知的表观遗传现象有：DNA 甲基化（DNA methylation）、基因组印记（genomic imprinting）、母体效应（maternal effects）、基因沉默（gene silencing）、核仁显性、休眠转座子激活和 RNA 编辑（RNA editing）等。

一、表观遗传学调控

LncRNA 招募染色质重构复合体到特定位点进而介导相关基因的表达沉默。例如，来源于 HOXC 基因座的 LncRNA HOTAIR，它能够招募染色质重构复合体 PRC2 并将其定位到 HOXD 位点，进而诱导 HOXD 位点的表观遗传学沉默。同样，Xist、Air、Kcnq1ot1 这些 LncRNA 都能够通过招募相应的重构复合体，利用其中的甲基转移酶如 Ezh2 或者 G9a 等实现表观遗传学沉默。

DNA 序列变异与那些改变基因表达水平、导致性状出现数量差异的因素和表观遗传状态密切相关。

人与人之间由于多种性状（如身高、头发颜色、行为和疾病易感性）的不同而有所差异。而遗传因素（先天性）和环境因素（后天性）是导致这种差异的“罪魁祸首”。近年来，大规模的遗传研究已经在人群中确定了成千上万个与不同性状密切相关的特殊 DNA 变异。然而，这些研究都没有回答一个关键问题：绝大多数 DNA 变异是通过何种方式来改变细胞行为、引起特殊性状（如身高）出现差异的。三篇文章（Kasowski 等撰写的文章，Kilpinen 等撰写的文章和 McVicker 等撰写的文章）向我们提供了一个研究框架，以便于研究者们探索人群中遗传变异与性状差异之间关联性的潜在机制。特别值得一提的是，他们发现 DNA 变异可通过序列特异性的转录因子活性，影响一种名为表观遗传学（epigenetics）的基因调控机制。

在过去的 10 年里，遗传学领域中最重要研究发现之一是：绝大部分与性状相关联的 DNA 变异都存在于一些曾经被称为“垃圾 DNA”的基因组区域中；这些基因组区域之所以被称为“垃圾 DNA”，是因为它们不能编码蛋白质。我们现在已经了解到，这些基因组区域含有一些调控性遗传因素，能够控制特定的基因在何时、何处、以何种程度进行表达，从而生成功能性的 RNA 与蛋白质产物。因此，大多数性状相关性 DNA 变异改变的并不是基因本身，而是改变了那些控制基因表达过程的调控性遗传因素。在过去的 3 年里，研究者们已经认识到一些基因调控性变异与性状（如血胆固醇浓度）和疾病（如糖尿病、骨关节炎、前列腺癌）密切相关。尽管我们获得了如此傲人的研究进展，但是大多数调节性变异究竟是如何改变基因表达的，我们还知之甚少。

众所周知，表观遗传学机制能够调控可遗传性的基因表达过程；但通常认为这种调控机制不依赖于 DNA 序列。DNA 被包裹在一种名为染色质的三维结构中，而染色质的基本重复单位是核小体（nucleosome），它的组成结构为一条长约 146 个核苷酸的 DNA，周围环绕着由名为