

## 长非编码 RNA 研究进展\*

陈晓敏 张栋栋 骆健俊 陈润生\*\*

(中国科学院生物物理研究所, 中国科学院核酸生物学重点实验室, 北京 100101)

**摘要** 长非编码 RNA 是指一类长度大于 200 个核苷酸、不编码蛋白质的非编码 RNA。越来越多的研究表明, 人类基因组中高达 90% 的非编码蛋白质的区段同样具有重要作用, 而不是所谓的“转录噪声”。针对长非编码 RNA 的功能研究表明, 其在转录起始的调控、转录及转录后的调控中均发挥着重要作用, 因而影响着各种各样的生物学过程。本综述围绕近几年长非编码 RNA 的研究成果, 总结了长非编码 RNA 的起源与进化、新型的长非编码 RNA 类型、典型的长非编码 RNA 作用机制以及长非编码 RNA 在发育与细胞重编程过程中的研究, 同时也概述了长非编码 RNA 与表观遗传调控和癌症的关系以及长非编码 RNA 研究的相关技术。系统发现长非编码 RNA 并阐明其功能机制, 将对现代生命科学具有重大的意义。

**关键词** 长非编码 RNA, 作用机制, 发育与细胞重编程, 表观遗传调控, 研究技术

**学科分类号** Q52

**DOI:** 10.3724/SP.J.1206.2014.00294

近年来非编码 RNA 的研究持续升温。从 1999 年开始到 2010 年的 12 年间有 6 年, 非编码 RNA 相关研究成果都入选美国《科学》(*Science*)杂志的年度十大科技突破。特别是 2010 年 12 月 17 日《科学》杂志在评选进入 21 世纪后第一个 10 年的十大科学突破时, 非编码领域被放在了第一位。2004 年, *Science* 以“隐蔽的 DNA 宝藏”为题, 指出在占人类基因组 90% 以上的所谓“无用 DNA”(junk DNA)序列中可能暗藏着大量的 DNA 调控元件、转座子和非编码 RNA 基因。2012 年, “人类 DNA 元件百科全书计划(ENCODE)”的最新结果表明<sup>[1]</sup>, 大约 80% 的 DNA 序列都能转录成 RNA。数目巨大、种类繁多的非编码 RNA 占细胞总 RNA 的绝大部分。这些非编码 RNA 无处不在, 而且参与了包括从干细胞维持、胚胎发育、细胞分化、凋亡、代谢、信号传导、感染以及免疫应答等几乎所有生理或病理过程的调控。由此可见, 非编码 RNA 的发现及其调控功能、机理的阐明, 对现代生命科学具有重大的意义。

由于非编码 RNA 种类繁多, 本文不再讨论已被学界广为了解的 rRNA, tRNA, 微 RNA (miRNA), Piwi 蛋白相作用的 RNA(piRNA), 小干扰 RNA(siRNA), 核内小 RNA(small nuclear RNA,

snRNA) 以及核仁小 RNA (small nucleolar RNA, snoRNA), 主要讨论长非编码 RNA(lncRNA)近年来的发展。国际上普遍认为的长非编码 RNA 是指长度大于 200 个核苷酸的 RNA 转录本, 以 200 个核苷酸划界是基于 RNA 的分离与提取程序<sup>[2]</sup>。我们认为如果以 100 个核苷酸划界可能与 RNA 的结构与功能联系更紧密, 短于 100 个核苷酸的 RNA 不易形成稳定的空间结构, 因此是依靠碱基配对为主行使功能的, 当 RNA 长于 100 个核苷酸时它们可能折叠成稳定的空间结构, 从而依靠生物大分子空间结构的相互作用发挥功能。在以前的工作中我们也曾提出过中等长度的非编码 RNA 的概念, 它们是指长度介于 50~500 个核苷酸的 RNA<sup>[3]</sup>。

长非编码 RNA 在转录水平参与了基因的表达调控, 在转录后水平参与了信使 RNA 的编辑与加工, 参与了翻译调控, 参与了基因组的重构, 参与了表观遗传调控、印迹调控和端粒系统的调控等,

\* 中国科学院战略先导项目(XDA01020402)和国家高技术研究发展计划(863)(2012AA020402 和 2012AA02A202)资助项目。

\*\* 通讯联系人。

Tel: 010-64888546, Fax: 010-64871293, E-mail: chenrs@sun5.ibp.ac.cn

收稿日期: 2014-10-08, 接受日期: 2014-10-11

其自身还可能成为核酶或核糖开关. 由于长非编码 RNA 多样的生物学功能, 科学家对其作用机制也提出了很多新的假说, 如: 长非编码 RNA 通过与蛋白质、RNA 和 DNA 相互作用形成各种模块来行使生物功能<sup>[4]</sup>; 长非编码 RNA 是细胞地址码的关键组分<sup>[5]</sup>; 非编码 RNA 是细胞内和细胞外的信号分子<sup>[6]</sup>等.

随着非编码 RNA 的不断发现, 国际上也出现了一些重要数据库来搜集相关信息为研究者使用, 如: NONCODE<sup>[7-8]</sup>、Y2K<sup>[9]</sup>等. 本文将就长非编码 RNA 近年来的发展做一综述.

## 1 长非编码 RNA 概述

### 1.1 非编码 RNA 研究历史

非编码 RNA 研究的发展源于科学界对基因组中非编码序列的认识(图 1). 早在 20 世纪 70 年代科学家们就注意到了非编码序列, 并将其称为“junk DNA”<sup>[10-11]</sup>. 其实, 当时科学家就猜测“junk DNA”并不是垃圾, 应当具有生物功能. 从 20 世纪 70 年代就发现了来自非编码序列的转录本, 如:

不均一核 RNA (heterogeneous nuclear RNA, hnRNA), 80 年代又发现了核内小 RNA (snRNA)和核仁小 RNA(snoRNA), 这说明基因组中的非编码序列是有信息发放的, 它们应当具有生物功能. 而真正对非编码序列及其转录产物的全面认识是始于 20 世纪 90 年代启动的“人类基因组计划”, 这一计划导致测定了很多物种的完整基因组序列, 从而了解了基因组中非编码序列的组成与结构<sup>[12]</sup>. 进入 21 世纪以后, 随着转录组研究的开展以及 ENCODE 计划的实施, 发现 75% 的人类基因组序列都有转录出的非编码 RNA, 这远比编码蛋白质的信使 RNA 多很多. 几年前, 日本的遗传研究所 (RIKEN) 在小鼠中获得了约 180 000 个全长的 RNA 转录本, 其中编码蛋白质的转录本仅有约 20 000 个, 其余约 160 000 个转录本全部归属于非编码 RNA<sup>[13]</sup>. 与此同时, 一些有代表性的功能非编码 RNA 分子(如 H19、Xist、lin-4、let-7、AIR)以及大量的 miRNAs、piRNAs 相继被发现. 这样, 一个崭新的、巨大的非编码核酸的世界就展现在了人们的面前.

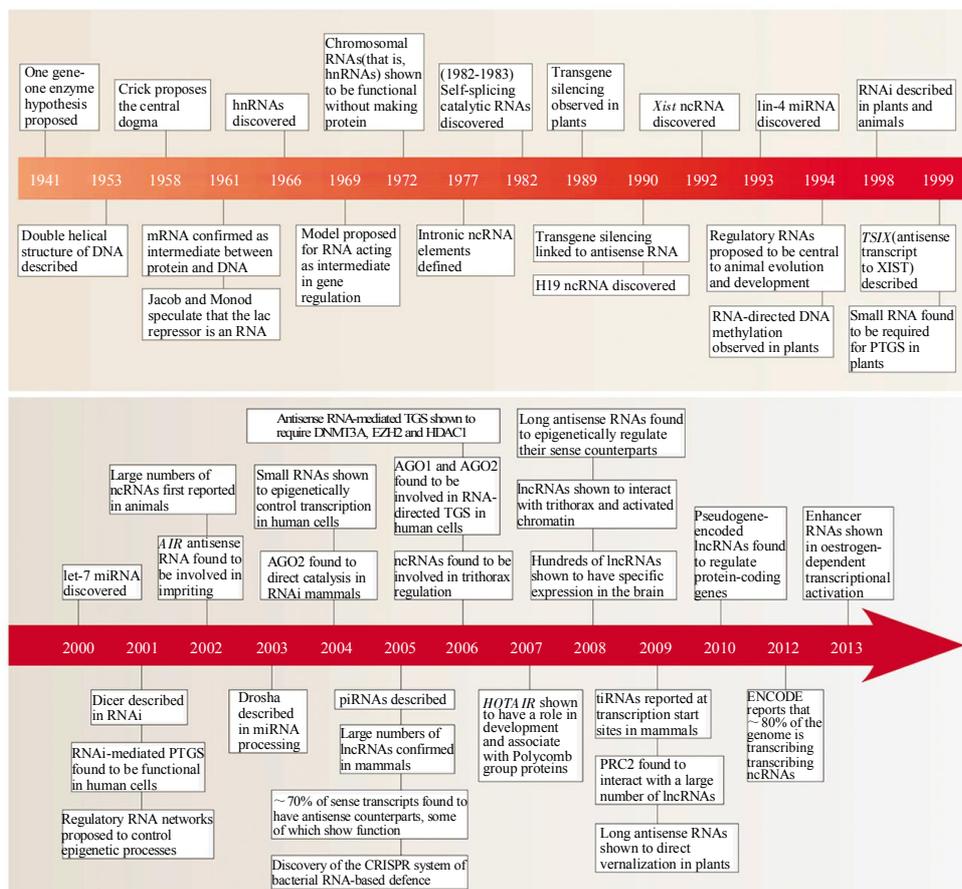


Fig. 1 The rise of regulatory RNA<sup>[2]</sup>

图 1 功能 RNA 的研究发展历程<sup>[2]</sup>

## 1.2 长非编码 RNA 的起源与进化

长非编码 RNA 是一种普遍的转录产物还是功能元件, 起初非常具有争议性, 它们在模式生物之间序列保守性差、表达量低, 导致被猜测可能是由于一些低保真性聚合酶产生的转录本, 而并不具有真正的功能<sup>[14]</sup>. 然而这种猜测被越来越多的深度测序分析所排除. 首先, 长非编码 RNA 的启动子区域以及剪切位点与蛋白编码基因具有一定相似性<sup>[15]</sup>. 其次, 尽管序列保守性相对 mRNA 较低, 长非编码 RNA 发挥作用可能并不是依赖于严格的序列上的保守性, 而是依靠二级空间结构<sup>[16]</sup>. 虽然目前确定的长非编码 RNA 大量涌现, 但绝大部分长非编码 RNA 在生命活动中的具体调控机制与功能模式仍需进一步的研究.

由于长非编码 RNA 的保守性相对较低, 目前普遍认为长非编码 RNA 的来源与进化可能存在以下 5 种机制: a. 蛋白编码基因的阅读框发生破坏而被转换成一个有功能的非编码 RNA(图 2a), 比如 Xist 的起源, Xist 的几个外显子与启动子区域被认为来源于蛋白编码基因 Lnx3 因转座子插入引起的阅读框突变. b. 来源于染色体重排, 两个不转录且相互远距离分离的序列区域发生并排而产生了一个多外显子的非编码 RNA(图 2b). 例如狗的一个非编码 RNA 就是来源于此种序列谱系演变(EST 序列 BM537447、C0597044 和 DN744681). c. 非编码基因可以通过反转录转座形成另一个有功能的

非编码逆转录基因, 或者另一个无功能的非编码逆转录假基因(图 2c). d. 有的非编码 RNA 含有的邻近重复序列可能来自于其中一个序列的两次随机复制(图 2d). e. 某个序列插入一个转座子(绿色部分)而形成有一个有功能的非编码 RNA(图 2e)<sup>[17]</sup>. 既然 RNA 在进化上是蛋白质的先驱, 长非编码 RNA 介导的转录调控很可能是一种古老的基因表达调控机制<sup>[18]</sup>.

研究发现, 中枢神经系统中长非编码 RNA 表达量与进化复杂性成正相关. 转录组分析显示, 脑中存在大量灵长类和人类特异表达的长非编码 RNA, 而这些特异性表达的长非编码 RNA 在脑发育过程中通常是瞬时表达的, 具有发育阶段的特异性<sup>[19-20]</sup>. 与编码基因相比, 长非编码 RNA 具有更快的进化速度, 如 HAR1A, 这些发现进一步证实长非编码 RNA 在脑的进化尤其是在脑认知和行为方面起到关键的调控作用<sup>[21-22]</sup>. 相反, 在脑中还存在一类从鸟类到哺乳动物都保守表达的长非编码 RNA, 他们具有相似的时空表达方式<sup>[23]</sup>. 这类长非编码 RNA 通常是由超保守区(ultraconserved regions, UCRs)的 DNA 转录, 与调控发育的关键基因有重叠或者是其互补序列. 这一类型的长非编码 RNA 可以作为分子支架募集特定的蛋白从而调控周围基因的表达<sup>[24]</sup>.

## 1.3 新型长非编码 RNA

长非编码 RNA 的发现与鉴定需要多种来源、类型的数据, 需结合 RNA-seq 数据、组蛋白 H3 第 4 位、第 36 位赖氨酸甲基化水平、转录起始位点、poly(A)位点等. 判定一新的转录本是否是非编码 RNA, 目前比较可行的办法也只能采用排除法去排除该转录本编码功能蛋白的可能. 真正的编码蛋白的基因通常具有以下几个特征, 可以当成区分 mRNA 与长非编码 RNA 的标准: a. 编码区域通常比随机期望的长度更长; b. 功能开放读码框中的核苷酸使用频率多使用非随机选取的编码子; c. 在进化进程中, 选择性压力倾向于编码序列中的核苷酸替换; d. 蛋白编码基因通常包含已知的蛋白结构域; e. 编码区域通常能找到与数据库中匹配的序列信息. 上述描述的蛋白编码潜力预测标准如果分开单独使用都会存在局限性. 将其中几项结合起来考虑就能比较好地从长非编码 RNA 中排除具蛋白编码能力的转录本<sup>[25]</sup>.

除了传统的长非编码 RNA 以外, 近几年来, 又发现了几类新类型的长非编码 RNA, 主要包括

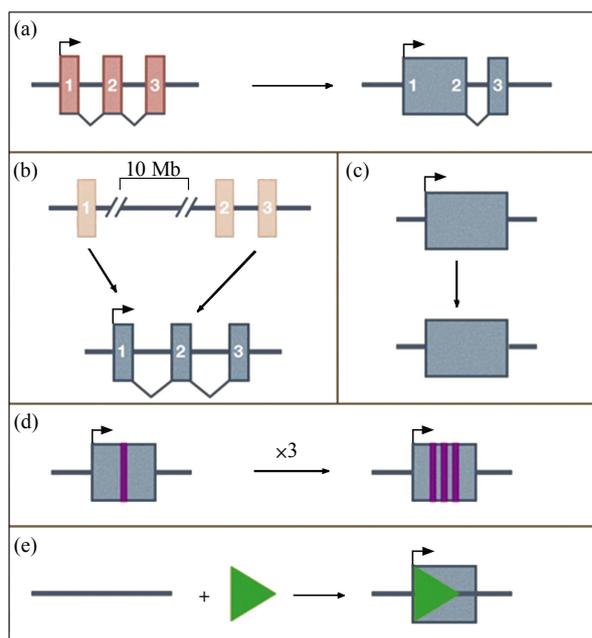


Fig. 2 Possible origins of lncRNAs<sup>[17]</sup>

图 2 长非编码 RNA 可能的起源类型<sup>[17]</sup>

增强子 RNA, 竞争性内源 RNA, 环形 RNA, 以及反向长非编码 RNA.

增强子 RNA(eRNA)是一类从增强子区域转录出的长度从几百碱基到数千碱基不等的非编码 RNA. 在小鼠神经元中, 已经检测到超过 12 000 个增强子, 其中有 2 000 个能结合 RNA 聚合酶 II, 并双向转录出长的且多数为不含 poly(A)尾巴的非编码 RNA<sup>[26]</sup>. 增强子 RNA 具有类似增强子的功能, 能够跨越染色体, 激活远端的启动子, 并在一定程度上调控编码基因的组织特异性表达. 在小鼠巨噬细胞中, 核受体转录因子 Rev-Erb 蛋白在转录调控区域的结合主要通过抑制增强子 RNA 的转录来抑制靶基因的表达<sup>[27]</sup>. 与 Rev-Erb 的抑制功能相反, 在人类乳腺癌细胞中, 雌激素受体(ERs)激活了基因表达, 不过和 Rev-Erb 一样, 它们主要也是通过结合增强子控制增强子 RNA 增高来发挥作用<sup>[28]</sup>. 最近的研究发现, 基因内部的增强子还有可能被作为可变的转录起始位点, 这些增强子区域在产生大量的、短的、双向的、没有 poly(A)尾巴的转录本的同时, 还可以转录出由多个外显子构成的、长的、带 poly(A)尾巴的增强子 RNA, 这种 RNA 也被称为多外显子增强子 RNA(meRNA)<sup>[29]</sup>. 通过调控增强子 RNA 的表达去影响受增强子 RNA 调控的靶基因的表达, 将是改变活细胞中基因表达的一条新途径.

近来, Pandolfi 等<sup>[30]</sup>提出了竞争性内源 RNA (ceRNA)调控基因表达的假说, 认为竞争性内源 RNA 活性能形成一种大规模转录调控网络, 扩大人类基因组中的功能性遗传信息, 并通过 miRNAs 应答元件, 作为 mRNAs 转录假基因, 以及长非编码 RNAs 相互“交流”的新语言. Pandolfi 等还认为 ceRNA 活性也在病理条件, 比如癌症中扮演了重要角色, 这对于解开一些癌症研究之谜具有非常重要的意义.

最近, 一些学者提出了环状 RNA 充当分子“海绵”, 结合并封闭了称作 miRNAs 的微小基因调控子的一种作用模式. 这一重要发现再次提醒人们: RNA 并不仅仅是 DNA 与编码蛋白之间的一个平凡信使. Hansen 等<sup>[31]</sup>发现一环状 RNA 的表达阻断了 miR-7, 它使得 miR-7 活性受到抑制, miR-7 靶基因表达增高, 并由此推测这是因为这一 RNA 环捕获和失活了 miR-7. 在斑马鱼中, 表达这一环状 RNA 或敲除 miR-7 可以改变大脑发育<sup>[32]</sup>. 研究表明, 环状 RNAs 也可能是细胞外 miRNA 的海

绵, 具有病毒 miRNAs 的结合位点, 从而破坏免疫应答.

反向长非编码 RNA(antisense long non-coding RNA)是指由已知的蛋白编码基因或者非编码基因的反义链转录的产物. 随着二代测序技术以及全基因组分析技术的发展, 发现人类基因组已注释的转录产物中 30%具有反向转录产物<sup>[33]</sup>. 反向长非编码 RNA 表达丰度都比较低, 一般是其正向转录本表达量的 1/10. 与蛋白编码基因相比, 反向长非编码 RNA 主要分布于细胞核. 其转录来源, 可能是与其正向转录本共同使用一个双向启动子, 也可能具有其独立的启动子, 发挥作用的方式有两种: 一种是其转录本身这个过程, 另一种是以成熟转录物发挥作用. 大家熟知的 Xist 和 ANRIL 是典型的反向长非编码 RNA.

随着单细胞测序技术以及生物信息学的发展, 加上 ENCODE 计划对于人类遗传信息的解读注释, 必将会对更多的占有人类基因组 98%的非编码转录产物有更深入的认识和功能发现.

## 2 长非编码 RNA 的功能

长非编码 RNA 是具有功能的转录本, 而不是转录噪声. 越来越多针对长非编码 RNA 的功能研究表明其在转录起始的调控、转录及转录后的调控中发挥着重要作用, 因而影响着各种各样的生物学过程, 比如, 剂量补偿、基因印迹、细胞周期、发育、配子形成等过程<sup>[34]</sup>. 通过募集染色质修饰复合物如多梳抑制复合物 II (PRC2), 长非编码 RNA 给非特异的酶活性提供了一种基因特异的靶向机制, 尽管多数功能已知的长非编码 RNA 抑制基因活性, 最近的一些研究包括来自 GENCODE 的数据都发现了许多长非编码 RNA 也可以激活基因的表达, 如增强子 RNA<sup>[35]</sup>. RNA 测序数据尤其是多物种间的长非编码 RNA 数据分析比较结合最近的整合研究显示, 长非编码 RNA 的功能并不单单取决于其分子大小、有无 poly(A)尾巴、剪切、转录方向甚至链特异性, 长非编码 RNA 在基因组中与靶基因的相对位置等结构信息也是影响其功能的一个不可忽视的基本重要因素<sup>[36]</sup>.

### 2.1 长非编码 RNA 实现生物功能的分子机制

越来越多的研究表明长非编码 RNA 的异常调控广泛参与了生物学的多种功能, 比如最早报道的与印迹相关的 Xist RNA, 以及 HOTAIR. Xist 基因的 5'端编码一个非编码 RNA-RepA, RepA 可以

与 PRC2 结合, 大量组蛋白被甲基化, 最终导致 X 染色体失活<sup>[37]</sup>. 此种 Xist RNA 介导的剂量补偿, 即基于 Xist RNA 表达的整条 X 染色体沉默机制, 平衡了雄性 XY 染色体和雌性 XX 染色体之间 X 连锁基因存在的剂量差别<sup>[38]</sup>. HOTAIR 是最早发现与肿瘤转移相关的长非编码 RNA 之一, 在原发和转移乳腺癌中均高表达, 高表达的 HOTAIR 与肿瘤侵袭、转移和患者预后不良密切相关. 典型的长非编码 RNA 作用分子机制可以概括如下<sup>[39]</sup>(图 3):

### 2.1.1 长非编码 RNA 作为信号分子

多数长非编码 RNA 由 RNA 聚合酶 II 转录产生, 长非编码 RNA 的表达具有细胞类型特异性并且受多种刺激调控, 暗示着他们的表达受转录调控. 此外, 长非编码 RNA 的转录具有很强的时空特异性, 表明其表达水平与发育及细胞环境紧密相关. 与蛋白质调控相比, 长非编码 RNA 信号分子因其转录剪切后即能折叠成高级结构, 不需要翻译过程, 因而能更快捷地行使调控功能(图 3 I).

### 2.1.2 长非编码 RNA 作为诱饵分子

非编码区域的增强子和启动子的活跃转录暗示着长非编码 RNA 在调节转录中的重要角色, 长非编码 RNA 通过结合目标蛋白或 miRNA 从而稀释了目标分子在细胞内的水平, 进而影响其功能(图 3 II). 如长非编码 RNA PANDA 在感应 DNA 损伤刺激后应激表达, 通过直接结合并拮抗核转录因子 NF-YA, 从而破坏依赖于 NF-YA 活性的

DNA 损伤所激活的凋亡通路<sup>[40]</sup>.

### 2.1.3 长非编码 RNA 的导向作用

通过与目标分子的结合, 长非编码 RNA 能指引核糖核蛋白复合体定位至特异的目标区域, 作用方式可以是顺式也可以是反式(图 3 III). 目前还无法简单地通过序列分析预测其会以何种方式发挥作用. 通过与 RNA 聚合酶作用以辅助转录的方式或者作为一些小的调节 RNA 分子的互补配对靶分子, 长非编码 RNA 能以反式作用的方式引导目标基因附近的染色质修饰状态改变. 通过与目标 DNA 分子结合形成 RNA:DNA 异源双链核酸分子, 或者 RNA:DNA:DNA 异源三链核酸分子, 或者 RNA 识别特异染色质的复合物表面特征, 长非编码 RNA 也能以反式的方式引导目标基因附近的染色质改变<sup>[39]</sup>.

### 2.1.4 长非编码 RNA 作为分子支架

不同于小 RNAs, 长非编码 RNA 因其结构及较长的核酸序列, 越来越多的研究已经表明其不同功能域可以结合不同的蛋白质复合体, 从而提供类似分子支架的功能, 以引导相关的大分子复合体在目标区域组装以协同发挥调控作用(图 3 IV). 如 HOTAIR, 由 HOXC 基因编码转录产生, 以分子支架的作用方式将两种不同的蛋白质复合物募集到染色体特定位点改变组蛋白甲基化修饰, 从而顺式调控并抑制 HOXD 基因的表达, 最终导致细胞侵袭转移能力的升高<sup>[41]</sup>.

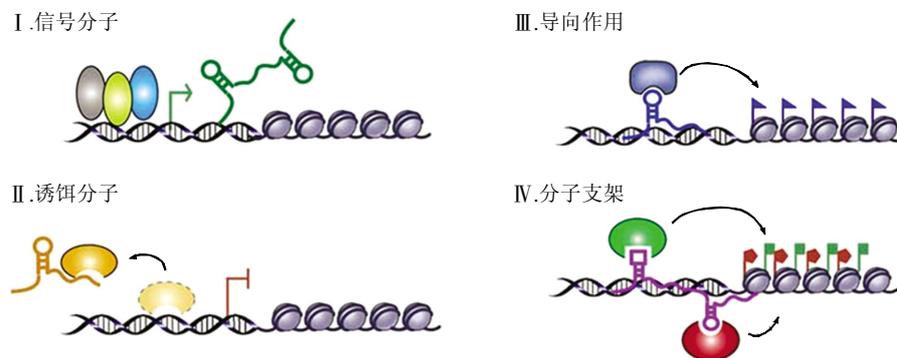


Fig. 3 Schematic diagram of the four archetypes of lncRNA mechanism<sup>[39]</sup>

图 3 长非编码 RNA 四种典型的分子作用方式示意图<sup>[39]</sup>

## 2.2 长非编码 RNA 与 microRNA 的相互作用

通过与其他生物大分子(DNA、RNA 以及蛋白质)发生相互作用, 非编码 RNA 能在多层面上影响

基因的表达调控, 包括染色质重塑、转录、mRNA 前体剪切、mRNA 周转、mRNA 翻译以及蛋白质的稳定性<sup>[42]</sup>. 过去 10 多年的研究也已经开始揭示

哺乳动物长非编码 RNA 与 microRNA (miRNA) 间存在着如下主要的四种形式的交互调节作用:

### 2.2.1 miRNA 介导的长非编码 RNA 降解

miRNA 能调节某些长非编码 RNA 的表达丰度, 从而影响目标长非编码 RNA 所参与的细胞生理、病理过程中的反应(图 4a). 例如 RNA 结合蛋白 HuR 能募集并促进 miRNA let-7 家族成员与 RISC 成分 Ago2 蛋白结合, 从而影响长非编码 RNA lincRNA-p21 以及 HOTAIR 的稳定性<sup>[43]</sup>.

### 2.2.2 长非编码 RNA 通过诱捕或 miRNA 海绵作用拮抗 miRNA 的功能

研究表明, 长非编码 RNA 也能影响 miRNA 的表达水平与功能, 如竞争性内源 RNA (ceRNA) 因其序列中存在类似于 miRNA 靶标 mRNA 的序列, 能诱捕 miRNA 与自身结合, 减少 miRNA 对目标 mRNA 的影响从而拮抗 miRNA 的功能(图 4b).

### 2.2.3 长非编码 RNA 与 miRNA 竞争性结合 mRNAs

除了以竞争性内源 RNA 的方式作用外, 长非编码 RNA 也会与 miRNA 竞争性的结合目标 mRNA(图 4c). 例如与 BACE1 部分序列反义互补的长非编码 RNA BACE1AS, 通过与 mRNA 的序列互补配对从而保护 mRNA 不受 miRNA 的结合, 避免由 miRNA-RISC 所介导的 RNA 降解<sup>[44]</sup>.

### 2.2.4 长非编码 RNA 作为 miRNA 产生的前体

研究表明, 有些长非编码 RNA 也能在成熟过程中由内含子或外显子区域序列剪切产生 miRNA (图 4d). 例如, linc-MD1 能产生 miR-206 和 miR-133b, 显示了 linc-MD1 在肌肉分化与肌营养不良中的额外的调控机制. 长非编码 RNA H19 也能产生 miR-675, 但此过程受 HuR 抑制<sup>[45]</sup>.

越来越多的证据已经表明长非编码 RNA 与 miRNA 在转录、转录后、翻译后多个层面调节着基因的表达. miRNA 对长非编码 RNA 的影响并非出乎意料, 毕竟长非编码 RNA 与 mRNA 在许多方面上都类似. 是否 miRNA 也调控着长非编码 RNA 的转录与剪切目前还无相关的报道, 同理, 长非编码 RNA 也可能在 miRNA 行使功能过程中发挥着尚无报道的作用, 比如是否能作为 miRNA 与靶 mRNA 相互作用的分子架, 此外, miRNA 与长非编码 RNA 也间接地发生着许多复杂的转录后调节机制以调节基因表达<sup>[46]</sup>.

## 3 长非编码 RNA 与发育

长非编码 RNA 在发育分化过程中发挥着非常重要的功能, 对多个组织以及不同细胞类型进行的深度测序以及深入的分析发现, 与编码基因相比, 长非编码 RNA 具有更强的细胞和组织特异性<sup>[47]</sup>; 此外, 在有机体分化的不同阶段, 长非编码 RNA 表达也存在着很大差别, 这也暗示了长非编码 RNA 可能是细胞命运决定的微调因子 (“fine-tuner”)<sup>[5, 47]</sup>. 研究证实, 长非编码 RNA 特异的时空表达与已知的细胞命运决定基因的表达具有很大的相关性, 可以通过正向和负向调控, 在分子水平决定细胞的分化方向<sup>[48-50]</sup>. 越来越多的研究表明, 长非编码 RNA 与神经、肌肉以及皮肤等分化过程紧密相关<sup>[51]</sup>.

### 3.1 长非编码 RNA 在中枢神经系统发育过程调控中的作用

长非编码 RNA 在建立和维持细胞特异性基因表达调控过程中发挥重要的作用, 而中枢神经系统是由神经元和神经胶质细胞组成, 是最复杂和多元化的组织, 因此研究中枢神经系统中长非编码 RNA 的表达具有重要意义. 在一项研究脑发育过程的工作中, Lim 等<sup>[20]</sup>分离了小鼠脑组织的 3 个不同区域 SVZ、OB 以及 DG 区, 通过 RNA-seq 和 ChIP-seq 进行分析, 发现超过 3 600 个差异表达的长非编码 RNA. 对长非编码 RNA 和 mRNA 进行

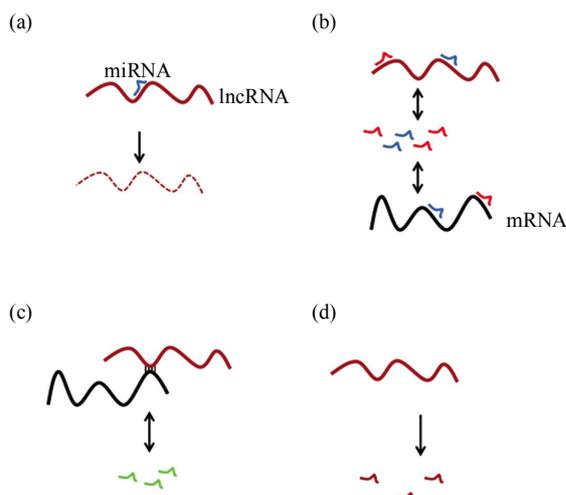


Fig. 4 Modes of direct post-transcriptional interaction among lincRNAs and miRNAs<sup>[46]</sup>

图 4 长非编码 RNA 与 miRNA 转录后相互作用方式<sup>[46]</sup>  
 (a) miRNA 介导长非编码 RNA 降解; (b) lincRNA 通过诱捕或 miRNA 海绵作用拮抗 miRNA; (c) lincRNA-miRNA 竞争性结合 mRNA; (d) lincRNA 作为 miRNA 前体.

聚类分析发现长非编码 RNA 比 mRNA 具有更高的组织特异性. 此外研究者通过 Capture-Seq 研究成人 SVZ 组织, 发现了超过 7 000 个特异表达的长非编码 RNA. 通过功能验证和生物信息学分析进一步筛选到两个重要的长非编码 RNA-Six3oc 和 *Dlx1as*. 敲降此两序列均会影响 SVZ 区第 7 天的分化, 其中敲降 Six3oc 导致 *Tuj1* 特异表达神经元细胞以及 OLIG2 阳性少突触细胞的分化阻滞, 而 *Dlx1as* 的缺失只影响 *Tuj1* 神经元细胞的分化. 在神经发生中, 长非编码 RNA 同样起到重要作用, 可以诱导神经系统的形成. 大规模筛选显示敲降很多长非编码 RNA 都会发生从胚胎干细胞向神经细胞分化的阻滞<sup>[52]</sup>.

中枢神经系统的复杂组成决定了其分化过程调控体系的复杂性, 长非编码 RNA 在多个层面可以发挥调控功能, 因此中枢神经系统的分化过程需要长非编码 RNA 的参与. 对长非编码 RNA 的进一步研究必将发现更多与神经系统发育分化以及退行性疾病相关的功能性长非编码 RNA.

### 3.2 长非编码 RNA 与心脏发育的相互关系

在组织发育过程中, 研究最清楚的是在中胚层发现并与小鼠心脏发育相关的两个长非编码 RNA——*Bvht* (brave heart) 和 *Fendrr* (Foxf1 adjacent non-coding developmental regulatory RNA). 研究发现, 在小鼠胚胎干细胞和新生小鼠心肌细胞中, 敲降 *Bvht* 会影响心脏特异性基因的表达, 从而影响心肌细胞的发育. 进一步的机制研究表明, *Bvht* 是与 PRC2 相互作用通过表观遗传学的修饰调控相关基因的表达. 但是, *Bvht* 只在小鼠中存在, 人和兔中没有表达, 也未发现功能同源序列<sup>[53]</sup>. 体外通过 RNAi 敲降 *Fendrr* 没有明显的表型, 但是胚胎细胞中完全敲除该基因则会由于心脏功能的损伤出现胚胎致死现象, 同时会影响体壁发育. 这一研究也从另一方面说明完全敲除模型对长非编码 RNA 研究的重要性. 进一步的机制研究发现, *Fendrr* 既可以与抑制信号 PRC2 复合物相互作用也可以与激活信号 MLL1 复合物相互作用, 在表观水平调控基因的表达. 但是与 *Bvht* 不同, *Fendrr* 在人类中存在同源序列, 发挥作用的机制类似, 可以与 PRC2 相互作用<sup>[54]</sup>.

### 3.3 长非编码 RNA 与骨骼肌发育的相互关系

在肌肉发生过程中, 第一个被发现的长非编码 RNA 是 *linc-MD1* (long non-coding RNA, muscle differentiation 1). 体外实验发现, MD1 在小鼠成

肌细胞向肌肉细胞分化的特定时间内瞬时表达. 机制研究表明, 该长非编码 RNA 可以作为增强子 RNA 控制肌肉细胞从早期到晚期的整个分化过程, *linc-MD1* 可以竞争性地与 miR-133 和 miR-135 结合, 从而保护转录因子 MAML1 和 MEF2C 的表达, 并进一步激活肌细胞的后期分化过程<sup>[49]</sup>.

*Linc-MD1* 在人、鼠中是序列保守的<sup>[55]</sup>, 但是在进行性肌营养不良患者体内其表达是显著下降的. 体外实验发现, 恢复这些肌营养不良细胞 *linc-MD1* 的表达, 可以修复肌细胞分化过程, 暗示着 *linc-MD1* 在肌肉分化过程中具有相对保守的功能<sup>[49]</sup>.

### 3.4 长非编码 RNA 与皮肤、造血以及脂肪发育的相互关系

长非编码 RNA 在上皮细胞分化中同样具有非常重要的功能. 对组织的分子生物学研究, 需要强有力的模型支撑. 而在整个发育生物学的研究中, 皮肤既是体内模型也便于体外研究, 为长非编码 RNA 的功能和机制研究提供了理想的模型. *Khavari* 等<sup>[56]</sup>在上皮细胞分化过程中发现两个重要的长非编码 RNA, ANCR 和 TINCR. ANCR 为长非编码 RNA 调控体壁干细胞分化提供了最早的证据, ANCR 的缺失会导致表皮干细胞分化的异常, 因此 ANCR 的主要功能是维持干细胞的干性, 阻止其向上皮细胞分化. 与 ANCR 不同, TINCR 促进干性细胞向上皮细胞分化, 祖细胞和分化的人类角质细胞转录组测序发现, TINCR 是分化过程中表达差异最大的长非编码 RNA. 研究发现, TINCR 缺陷的上皮细胞缺少终末分化的超微结构, 包括透明角质颗粒和完整的层状颗粒. 进一步对其作用机制进行研究发现, TINCR 可以与 STAU1 蛋白结合形成 TINCR-STAU1 复合物, 但是该复合物与 mRNA 结合从而稳定分化相关的 mRNA, 比如 KRT80 等蛋白, 从而保证细胞分化的顺利进行<sup>[57]</sup>.

文献报道, 在造血细胞分化过程中也有长非编码 RNA 参与, 其中 *lincRNA-EPS* 是在小鼠红细胞分化过程中发现的长非编码 RNA. 在小鼠造血祖细胞中敲降 *lincRNA-EPS* 会阻滞其分化过程引起凋亡. 但是该长非编码 RNA 具体的作用机制还不明确<sup>[58-59]</sup>. 近年来我们实验室研究了间充质干细胞 (MSC) 向脂肪分化过程中差异表达的长非编码 RNA, 发现了一条有意义的序列, 该序列可以通过表观修饰的改变阻滞 MSC 的成脂过程(未发表结果).

### 3.5 长非编码 RNA 与细胞重编程

细胞重编程是指分化的细胞在特定条件下被逆转后恢复到全能型或多能性的状态, 或者形成胚胎干细胞的过程. 由成纤维细胞或者其他成体细胞向 iPS 细胞转化的细胞重编程过程证明了细胞命运的可塑性. 在整个重编程过程中, 表观基因组发生了完全的重塑, 基因表达谱也发生巨大的变化, 从而产生大量差异表达的 mRNA、miRNA 和长非编码 RNA<sup>[60-62]</sup>. 目前对细胞重编程过程尤其是 iPS 过程中长非编码 RNA 的研究还非常有限, 可用数据相对较少, 因此在重编程过程中对长非编码 RNA 的研究大部分与胚胎干细胞相结合. 最新的研究发现, 在重编程过程以及小鼠和人类的胚胎干细胞中存在特异性的长非编码 RNA, 这些长非编码 RNAs 与重编程重要因子 OCT4、NANOG 以及 SOX2 表达有很大相关性<sup>[63]</sup>. 进一步生物信息学的分析显示, 这些长非编码 RNA 的启动子区至少有一个重编程重要因子的结合位点<sup>[64]</sup>.

对胚胎干细胞转录组进一步分析, 发现大量的长非编码 RNA 参与细胞多能性分化的调控过程. Guttman 等<sup>[64]</sup>针对其中的 226 个序列设计 shRNA, 通过功能性缺失实验寻找与干细胞干性维持以及抑制分化相关的长非编码 RNA. 研究发现, 其中的 26 个长非编码 RNA 敲降后会影响 Nanog 的表达, 从而导致 ES 细胞的多能性丧失. 另有 30 个长非编码 RNA 敲降之后会导致胚胎干细胞向特定的谱系分化. Rinn 等<sup>[65]</sup>在细胞重编程过程中发现一条重要长非编码 RNA-lincRNA-RoR, 研究发现敲降或过表达该序列均会显著影响成纤维细胞的重编程过程. 与 Guttman 之前报道的结果不同, RNA pull-down 实验表明 lincRNA-RoR 可以与 miR145-5p、181a-5p、99b-3p 以及他们的靶基因 Ago2 结合. 已有研究表明, 这些 miRNAs 能够调控细胞重编程过程的核心分子包括 Pou5f1、Sox2 和 NANOG, 因此, lincRNA-RoR 作为竞争性内源 RNA 在细胞重编程过程中发挥了重要的作用.

重编程过程中长非编码 RNA 的研究还处在相对早期的阶段, 分离到的功能性长非编码 RNA 相对有限, 具体的作用机制还不清楚, 因此这是一个很有潜力的发展领域. 胚胎干细胞的非编码转录组现在得到大家更多的关注, 同时在胚胎干细胞中存在大量的 RNA-seq 和染色质修饰 ChIP-seq 的数据, 为长非编码 RNA 功能的预测提供了很大的帮助. 我们相信随着研究的深入, 会有更多功能明确、机

制清晰的更重要的长非编码 RNA 被发现, 从而进一步加深对细胞重编程过程的认识.

## 4 长非编码 RNA 与表观遗传调控

长非编码 RNA 能作用于单个目标基因或一簇目标基因家族, 对其在转录调控上的作用已经有广泛的研究, 研究也发现许多长非编码 RNA 与蛋白质发生作用以调节转录激活或沉默. 一些长非编码 RNA 已被证实可在转录后调控基因表达, 例如长非编码 RNA MALAT1, 能抑制丝氨酸/精氨酸蛋白活性, 从而调节 mRNA 前体的可变剪切<sup>[66]</sup>.

最近的研究结果显示, 长非编码 RNA 在组蛋白修饰、DNA 甲基化等表观遗传修饰过程中也发挥重要的作用, 通过结合并募集特定的表观修饰酶复合物至目标基因区域, 改变靶基因染色质或 DNA 修饰状态从而影响该靶基因的表达<sup>[41, 67-68]</sup>. 如长非编码 RNA HOTAIR, 它的 5'端能够结合 PRC2 复合物, 3'端结合 LSD1/CoREST/REST 复合物, 并将整个复合物锚定在特定的基因区域, 使该区域的组蛋白 H3 的第 27 位赖氨酸甲基化, 而使第 4 位的赖氨酸去甲基化, 通过协调这种关系来调控基因表达(图 5)<sup>[41, 69]</sup>.

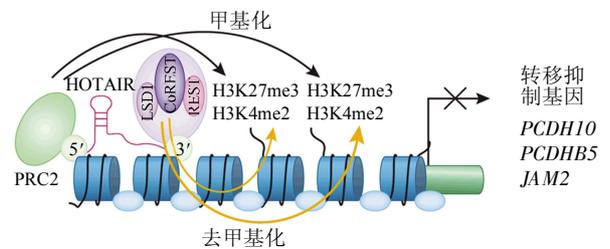


Fig. 5 Schematic presentation of HOTAIR function in breast cancer progression<sup>[70]</sup>

图 5 HOTAIR 在乳腺癌进展中的功能示意图<sup>[70]</sup>

研究也发现, 长非编码 RNA ecCEBP (extra-coding CEBPA) 的转录本表达起始于 CEBPA 上游 2 kb, 同向转录, 但转录产物不含 poly(A) 且多富集在细胞核内, 通过与 DNA 甲基化酶 DNMT1 结合而减弱 CEBPA 启动子区域的甲基化从而促进 CEBPA 基因的表达<sup>[68]</sup>. 肿瘤抑制因子 TCF21 的一个反义长非编码 RNA TARID, 通过与 TCF21 启动子以及 GADD45A 相互作用, 从而募集了 DNA 去甲基化酶 TET 蛋白至 TCF21 启动子区, 表明长非编码 RNA 能介导特异目标基因的 DNA 去甲基化

过程从而激活靶基因的表达<sup>[67]</sup>。

起初, 人们认为长非编码 RNA 的功能是在其位置周围发挥作用的, 比如, 调控邻近基因的表达等。然而, 最近研究认为, 长非编码 RNA 能够和上千个不同位置的染色质有相互作用, 并进而大规模地调控基因表达。进一步对这些结合蛋白分析发现, 功能性长非编码 RNA 可以结合各种染色质修饰复合物包括“阅读器”(readers)(Prc1、Cbx1 与 Cbx3)、“书写器”(writers)(Tip60/P400、Prc2、Setd8、Eset 与 Suv39h1)和“擦除器”(erasers)(Jarid1b、Jarid1c 与 Hdac1), 以及 DNA 甲基化酶或去甲基化酶<sup>[67-68]</sup>。然而长非编码 RNA 结合远程的结合位点的分子机制目前还不是很清楚<sup>[71]</sup>, 还有待进一步深入地研究。

## 5 长非编码 RNA 和癌症

长非编码 RNA 与癌症等有着密切的关系, 对癌症的发生、发展及转移产生重要的影响。长非编

码 RNA 在癌症中是至关重要的, 但是我们对其功能的认识还很浅。有一些长非编码 RNA, 比如 PCA3、PCGEM1、PCAT1 等, 是高度在前列腺癌中特异表达的非编码 RNA, 可以作为有效的生物标记物<sup>[72]</sup>。

近年来的研究表明, 有多种长非编码 RNA 在原发性肝癌中表达水平发生了显著变化, 并具有重要作用。H19 是第一个被发现的非编码 RNA 基因, 在肝脏中, H19 能与血管生成素和成纤维细胞生长因子相互作用, 可改变其表达而诱发肿瘤<sup>[73]</sup>。HULC 是在肝癌中上调表达最高的转录本, 也是首个被发现在肝癌胞质中过表达的长非编码 RNA<sup>[74]</sup>。在肝癌细胞中, 位于核心启动子区的 cAMP 反应元件结合蛋白结合位点及 PKA 途径在 HULC 上调中具有重要作用。另一方面, HUCL 能下调包括 miR-372 在内的一系列 miRNA, 从而在肝癌中发挥重要作用<sup>[75]</sup>。和肝癌相关的长非编码 RNA 如表 1 所示<sup>[76]</sup>。

Table 1 Examples of lncRNA associated with hepatocellular carcinoma<sup>[76]</sup>  
表 1 肝癌相关的长非编码 RNA 列表<sup>[76]</sup>

长非编码 RNA	染色体上的位置	大小(kb)	肝癌中的可能作用
HULC	Chr6	0.5	肝癌细胞中表达上调, 高表达常与组织学分级及乙肝病毒相关
TUC338	Chr12	0.59	肝硬化组织以及肝癌细胞中表达增加调节细胞生长速度
UCA1/CUDR	Chr19	1.4~27	具有化疗耐受性
HEIH	Chr5	1.7	与乙肝病毒相关
MEG3	Chr14	~1.8	肝癌细胞中表达下调, 与甲基化相关检测治疗效的潜在分子标记物
HTOAIR	Chr12	2.3	抑制表达引起细胞侵袭能力降低化学感性增加
HOTTIP	Chr7	7.9	肝癌细胞中表达上调, 预测肝癌进展
MALAT-1	Chr11	8.7	肝癌细胞中表达上调, 与癌症转移复发相关
LINC-ROR	Chr18	22.8	低氧状态下肿瘤细胞的存活相关

肺癌转移相关的转录物 MALAT1 (metastasis associated lung adenocarcinoma transcript 1), 是一个长度大于 8 000 个核苷酸的长非编码, 主要在细胞核内发挥作用, 它在多种物种中具有保守性。研究表明 MALAT1 在肺和胰腺中高表达<sup>[77]</sup>, 通过 shRNA 的方法敲降 MALAT1, 显示 MALAT1 通过调节包括 caspase-3, -8, Bax, Bcl-2, BclxL 等基因的表达, 从而影响细胞的生长, 细胞周期以及细胞侵袭, 功能试验表明, MALAT1 的 3'端是其发挥作用的关键区域<sup>[78]</sup>。因此, 有望通过对 MALAT1 的 3'端的干扰进行靶向治疗。

胰腺癌基因表达标志 PCGEM1 (prostate cancer

gene expression marker1)是一个胰腺癌相关的长非编码 RNA, 在胰腺癌早期的发生和发展中起重要作用, PCGEM1 基因长度为 1 603 个核苷酸, 在胰腺癌细胞和 NIH3T3 细胞中过表达 PCGEM1, 会促进细胞增殖和克隆形成能力, 提示 PCGEM1 是一个致癌基因。PCGEM1 的表达与胰腺癌细胞发展到去势复发阶段有重要相关性<sup>[79]</sup>, 去势复发是病发的最后一个阶段, 对于患病者至关重要的意义。

除了以上列举的一些癌症相关的长非编码 RNA, 还有一些诸如 aHIF、ANRIL、Oct4-pg、PTENP1 和 BC200 等在神经母细胞瘤、乳腺癌、胶质瘤、结直肠癌、神经退行性疾病中有功能的

长非编码 RNA。但是由于其自身结构的复杂性，对于它们如何发挥作用的机制还需进一步深入地研究。近年来，长非编码 RNA 对于疾病发生发展的科学价值和临床意义引起人们越来越多的兴趣，它不仅可以为包括癌症在内的许多复杂疾病的诊断和治疗提供新的依据和靶点，而且有助于人们进一步认识高等真核生物极其复杂的调控网络。

## 6 长非编码 RNA 研究的相关技术

对长非编码 RNA 的研究直到最近几年才得到广泛的关注，与长非编码 RNA 相关的数据分析及实验技术也随之发展起来。与蛋白质研究方法不同，长非编码 RNA 的稳定性相对较差，容易降解。此外，其表达丰度也远低于 mRNA。与 miRNA 相比，长非编码 RNA 因其长度更长，存在二级结构，发挥功能的方式也更复杂。因此，对长非编码 RNA 的研究除去常规的方法外还有很多是其特有的研究方法。

### 6.1 长非编码 RNA 的发现和系统鉴定

长非编码 RNA 的发现依赖于全基因组范围内对转录本的筛选，主要是注释为非编码的序列，比如缺少有效的开放阅读框等。现在专门针对长非编码 RNA 的数据库还比较少，NONCODE 是现今收录非编码 RNA 最多，注释最全的专门用于非编码 RNA 查询的数据库<sup>[7-8]</sup>。在数据分析的基础上可以进一步通过实验手段以系统鉴定长非编码 RNA。利用长非编码 RNA 检测微阵列芯片，可以在芯片包含的探针信息的基础上筛选差异表达的长非编码 RNA<sup>[9]</sup>，快速、高效，但是其局限性在于所发现新的差异表达的长非编码 RNA 受限于探针的预设计。利用 RNA-seq 开展的转录组测序是现在分析差异表达长非编码 RNA 的另一有效途径。通过测序可以发现现有数据库中没有注释的序列，这是其独特的优势。

### 6.2 长非编码 RNA 序列信息的确定以及功能和机制研究

现在数据库中大部分长非编码 RNA 信息都是通过测序发现并注释，因此其具体的序列信息并不是很明确，需要通过实验进一步鉴定其序列信息。常规的可以通过 RACE 以确定其两端序列，通过 Northern 杂交检测整条序列的长度以及主要的转录本，最终结合已有的测序信息以确定目标长非编码 RNA 的序列信息。利用 RT-qPCR 可以定量地分析样本中目标序列的表达水平，以进一步验证芯片和

测序结果。

对长非编码 RNA 功能的研究与传统对基因的研究相似，可以通过敲降和过表达目标序列以发现具体的功能，研究方法包括 RNAi、质粒和慢病毒转染以及 TALEN/CRISPR 等技术。对于长非编码 RNA 作用机制的研究，RNA pull-down 联合蛋白质谱分析是最常用的寻找与长非编码 RNA 相互作用蛋白的方法。此外，RNA 结合蛋白免疫沉淀 (RNA-binding protein immunoprecipitation, RIP) 联合 qPCR 实验可以用来进一步验证两者的相互作用<sup>[4]</sup>。

长非编码 RNA 不仅与蛋白质相互作用也可与 DNA 相互作用，RNA 纯化的染色质分离 (chromatin isolation by RNA purification, ChIRP) 技术被广泛应用。首先用戊二醛固定细胞，以维持长非编码 RNA 与染色质的相互作用，然后进行细胞裂解和超声破碎，接着用生物素标记的寡核苷酸探针与靶长非编码 RNA 杂交，基于生物素和链霉亲和素相互作用的原理，用链霉亲和素磁珠来分离、纯化染色质复合体，最后从纯化的染色质复合体中分离蛋白质、RNA 或 DNA 以进行下游的分析。Chang 等<sup>[10]</sup>发展并完善了这一技术，与高通量测序结合用以在全基因组范围内扫描长非编码 RNA 的结合位点。

随着技术水平的飞跃发展，针对长非编码 RNA 研究的新技术必将会不断涌现，从而为长非编码 RNA 研究的快速发展提供必要的技术保障。

## 7 展 望

近年来，对非编码 RNA 的研究使我们对一些经典的概念产生了新的认识，令我们不得不重新考虑遗传学和分子生物学的最核心问题：什么是基因？而“中心法则”等一批近乎经典的观念，在基因组产生的大量非编码转录本以及迅速涌现的非传统实验现象面前，正在经受着冲击。同时，也为我们创造新的研究成果提供了巨大机会。

## 参 考 文 献

- [1] Consortium E P. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, **489**(7414): 57-74
- [2] Morris K V, Mattick J S. The rise of regulatory RNA. *Nat Rev Genet*, 2014, **15**(6): 423-437
- [3] Deng W, Zhu X P, Skogerbo G, *et al.* Organization of the *Caenorhabditis elegans* small non-coding transcriptome: Genomic features, biogenesis, and expression. *Genome Research*, 2006,

- 16(1): 20–29
- [4] Guttman M, Rinn J L. Modular regulatory principles of large non-coding RNAs. *Nature*, 2012, **482**(7385): 339–346
- [5] Batista P J, Chang H Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell*, 2013, **152**(6): 1298–1307
- [6] Dinger M E, Mercer T R, Mattick J S. RNAs as extracellular signaling molecules. *J Mol Endocrinol*, 2008, **40**(4): 151–159
- [7] Liu C N, Bai B Y, Skogerbo G, *et al.* NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Research*, 2005, **33**: D112–D115
- [8] Xie C, Yuan J, Li H, *et al.* NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*, 2014, **42**(Database issue): D98–103
- [9] Erdmann V A, Szymanski M, Hochberg A, *et al.* Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res*, 2000, **28**(1): 197–200
- [10] Comings D E. The structure and function of chromatin. *Adv Hum Genet*, 1972, **3**: 237–431
- [11] Ohno S. So much "junk" DNA in our genome. *Brookhaven Symp Biol*, 1972, **23**: 366–370
- [12] Kung J T, Colognori D, Lee J T. Long noncoding RNAs: past, present, and future. *Genetics*, 2013, **193**(3): 651–669
- [13] Carninci P, Kasukawa T, Katayama S, *et al.* The transcriptional landscape of the mammalian genome. *Science*, 2005, **309**(5740): 1559–1563
- [14] Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol*, 2007, **14**(2): 103–105
- [15] Ponjavic J, Ponting C P, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*, 2007, **17**(5): 556–565
- [16] Ulitsky I, Shkumatava A, Jan C H, *et al.* Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, 2011, **147**(7): 1537–1550
- [17] Ponting C P, Oliver P L, Reik W. Evolution and Functions of Long Noncoding RNAs. *Cell*, 2009, **136**(4): 629–641
- [18] Orom U A, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*, 2013, **154**(6): 1190–1193
- [19] Mercer T R, Dinger M E, Sunken S M, *et al.* Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA*, 2008, **105**(2): 716–721
- [20] Ramos A D, Diaz A, Nellore A, *et al.* Integration of genome-wide approaches identifies lincRNAs of adult neural stem cells and their progeny *in vivo*. *Cell Stem Cell*, 2013, **12**(5): 616–628
- [21] Pollard K S, Salama S R, Lambert N, *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 2006, **443**(7108): 167–172
- [22] Lindblad-Toh K, Garber M, Zuk O, *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 2011, **478**(7370): 476–482
- [23] Dorus S, Vallender E J, Evans P D, *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell*, 2004, **119**(7): 1027–1040
- [24] Amaral P P, Neyt C, Wilkins S J, *et al.* Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development. *RNA*, 2009, **15**(11): 2013–2027
- [25] Ulitsky I, Bartel D P. lincRNAs: genomics, evolution, and mechanisms. *Cell*, 2013, **154**(1): 26–46
- [26] Kim T K, Hemberg M, Gray J M, *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 2010, **465** (7295): 182–187
- [27] Lam M T, Cho H, Lesch H P, *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 2013, **498**(7455): 511–515
- [28] Li W, Notani D, Ma Q, *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 2013, **498**(7455): 516–520
- [29] Natoli G, Andrau J C. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet*, 2012, **46**: 1–19
- [30] Salmena L, Poliseno L, Tay Y, *et al.* A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. *Cell*, 2011, **146** (3): 353–358
- [31] Hansen T B, Jensen T I, Clausen B H, *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013 (doi: 10.1038/nature11993.)
- [32] Memczak S, Jens M, Elefsinioti A, *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 2013 (doi: 10.1038/nature11928.)
- [33] Djebali S, Davis C A, Merkel A, *et al.* Landscape of transcription in human cells. *Nature*, 2012, **489**(7414): 101–108
- [34] Yang L, Froberg J E, Lee J T. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci*, 2014, **39**(1): 35–43
- [35] Flynn R A, Chang H Y. Active chromatin and noncoding RNAs: an intimate relationship. *Curr Opin Genet Dev*, 2012, **22**(2): 172–178
- [36] Ernst C, Morton C C. Identification and function of long non-coding RNA. *Front Cell Neurosci*, 2013, **7**: 168
- [37] Zhao J, Sun B K, Erwin J A, *et al.* Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, 2008, **322**(5902): 750–756
- [38] Nesterova T B, Slobodyanyuk S Y, Elisaphenko E A, *et al.* Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res*, 2001, **11**(5): 833–849
- [39] Wang K C, Chang H Y. Molecular mechanisms of long noncoding RNAs. *Mol Cell*, 2011, **43**(6): 904–914
- [40] Hung T, Wang Y, Lin M F, *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*, 2011, **43**(7): 621–629
- [41] Tsai M C, Manor O, Wan Y, *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 2010, **329**(5992): 689–693
- [42] Mercer T R, Mattick J S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol*, 2013, **20**(3):

- 300–307
- [43] Yoon J H, Abdelmohsen K, Srikantan S, *et al.* LincRNA-p21 suppresses target mRNA translation. *Mol Cell*, 2012, **47** (4): 648–655
- [44] Faghihi M A, Zhang M, Huang J, *et al.* Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol*, 2010, **11**(5): R56
- [45] Keniry A, Oxley D, Monnier P, *et al.* The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol*, 2012, **14**(7): 659–665
- [46] Yoon J H, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol*, 2014, **34C**: 9–14
- [47] Rinn J L, Chang H Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*, 2012, **81**: 145–166
- [48] Ebert M S, Sharp P A. Roles for microRNAs in conferring robustness to biological processes. *Cell*, 2012, **149**(3): 515–524
- [49] Cesana M, Cacchiarelli D, Legnini I, *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, 2011, **147**(2): 358–369
- [50] Wang Y, Xu Z, Jiang J, *et al.* Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Developmental Cell*, 2013, **25**(1): 69–80
- [51] Flynn R A, Chang H Y. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell*, 2014, **14** (6): 752–761
- [52] Ng S Y, Johnson R, Stanton L W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J*, 2012, **31**(3): 522–533
- [53] Klattenhoff C A, Scheuermann J C, Surface L E, *et al.* Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, 2013, **152**(3): 570–583
- [54] Grote P, Wittler L, Hendrix D, *et al.* The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental Cell*, 2013, **24**(2): 206–214
- [55] Twayana S, Legnini I, Cesana M, *et al.* Biogenesis and function of non-coding RNAs in muscle differentiation and in Duchenne muscular dystrophy. *Biochem Soc Trans*, 2013, **41**(4): 844–849
- [56] Kretz M, Webster D E, Flockhart R J, *et al.* Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev*, 2012, **26**(4): 338–343
- [57] Kretz M, Siprashvili Z, Chu C, *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*, 2013, **493**(7431): 231–235
- [58] Hu W, Yuan B, Flygare J, *et al.* Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev*, 2011, **25**(24): 2573–2578
- [59] Sun L, Goff L A, Trapnell C, *et al.* Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci USA*, 2013, **110**(9): 3387–3392
- [60] Chin M H, Mason M J, Xie W, *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell*, 2009, **5**(1): 111–123
- [61] Ball M P, Li J B, Gao Y, *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*, 2009, **27**(4): 361–368
- [62] Doi A, Park I H, Wen B, *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet*, 2009, **41**(12): 1350–1353
- [63] Dinger M E, Amaral P P, Mercer T R, *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*, 2008, **18**(9): 1433–1445
- [64] Guttman M, Donaghey J, Carey B W, *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 2011, **477**(7364): 295–300
- [65] Loewer S, Cabili M N, Guttman M, *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*, 2010, **42**(12): 1113–1117
- [66] Tripathi V, Ellis J D, Shen Z, *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, 2010, **39**(6): 925–938
- [67] Arab K, Park Y J, Lindroth A M, *et al.* Long Noncoding RNA TARID Directs Demethylation and Activation of the Tumor Suppressor TCF21 *via* GADD45A. *Mol Cell*, 2014, **55**(4): 604–614
- [68] Di Ruscio A, Ebralidze A K, Benoukrif T, *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*, 2013, **503**(7476): 371–376
- [69] Katayama S, Tomaru Y, Kasukawa T, *et al.* Antisense transcription in the mammalian transcriptome. *Science*, 2005, **309**(5740): 1564–1566
- [70] Croce C M. LINCing chromatin remodeling to metastasis. *Nat Biotechnol*, 2010, **28**(9): 931–932
- [71] Vance K W, Ponting C P. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet*, 2014, **30**(8): 348–355
- [72] Walsh A L, Tuzova A V, Bolton E M, *et al.* Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'?. *Trends Mol Med*, 2014, **20**(8): 428–436
- [73] Court F, Baniol M, Hagege H, *et al.* Long-range chromatin interactions at the mouse Igf2/H19 locus reveal a novel paternally expressed long non-coding RNA. *Nucleic Acids Res*, 2011, **39**(14): 5893–5906
- [74] Panzitt K, Tschernatsch M M, Guelly C, *et al.* Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology*, 2007, **132** (1): 330–342
- [75] Wang J, Liu X, Wu H, *et al.* CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res*, 2010, **38**(16): 5366–5383
- [76] Takahashi K, Yan I, Haga H, *et al.* Long noncoding RNA in liver diseases. *Hepatology*, 2014, **60**(2): 744–753

- [77] Ji P, Diederichs S, Wang W, *et al.* MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 2003, **22** (39): 8031–8041
- [78] Xu C, Yang M, Tian J, *et al.* MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. *Int J Oncol*, 2011, **39**(1): 169–175
- [79] Romanuik T L, Wang G, Morozova O, *et al.* LNCaP Atlas: gene expression associated with *in vivo* progression to castration-recurrent prostate cancer. *BMC Med Genomics*, 2010, **3**: 43
- [80] Li J, Chen Z, Tian L, *et al.* LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut*, 2014
- [81] Chu C, Qu K, Zhong F L, *et al.* Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*, 2011, **44**(4): 667–678

## Research Progress on Long Noncoding RNAs\*

CHEN Xiao-Min, ZHANG Dong-Dong, LUO Jian-Jun, CHEN Run-Sheng\*\*

(Chinese Academy of Sciences Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract** Long noncoding RNAs are non-protein coding transcripts greater than 200 bp in length. Rapidly growing studies showed that these 90% non-protein coding regions in human genome, instead of being "transcription noise", were also playing important roles on transcription initiation, elongation and post-transcriptional regulation *etc.*, thus affecting various biological processes. Based on recent research progress on long noncoding RNAs, this review aims to summarize their origin and evolution, novel kinds of long noncoding RNAs, typical functional mechanisms of long noncoding RNAs, and their roles in development and reprogramming. We also tried to update studies on epigenetic regulation of long noncoding RNAs as well as their roles in cancer research, and some closely related modern technologies for long noncoding RNAs. Systematic screen of long noncoding RNAs and the elaboration of their functional mechanisms will play critical roles in the progress of modern life sciences.

**Key words** long noncoding RNA, mechanism, development and reprogramming, epigenetic regulation, technologies

**DOI:** 10.3724/SP.J.1206.2014.00294

---

\* This work was supported by grants from Priority Research Program of The Chinese Academy of Sciences (XDA01020402) and Hi-Tech Research and Development Program of China (2012AA020402, 2012AA02A202).

\*\*Corresponding author.

Tel: 86-10-64888546, Fax: 86-10-64871293, E-mail: chenrs@sun5.ibp.ac.cn

Received: October 8, 2014

Accepted: October 11, 2014