

DOI: 10.13376/j.cblls/2016089

文章编号: 1004-0374(2016)06-0695-08



崔恒宓, 博士, 1997年获瑞典 Uppsala 大学分子与发育遗传学博士学位, 其后任美国约翰霍普金斯大学博士后、讲师和助理教授, 美国国家卫生研究院 (NIH) 资深科学家。2013 年被扬州大学作为重点人才引进, 聘为扬州大学特聘教授、表观遗传学及表观基因组学研究所所长, 江苏省高层次双创人才。在国外从事生物科学前沿研究 20 多年, 在 SCI 杂志发表论文 60 余篇, 包括 *Nature*、*Science*、*Nature Medicine*、*Nature Genetics*、*Cancer Cell* 等世界一流杂志, 引用超过 6 000 次。现兼任 *BMC Medical Genetics* 副主编 (Associate Editor) 和 *The Scientific World Journal* 等多个国际杂志编委。目前, 实验室的研究兴趣在于人类疾病的表观遗传学机制、内源性反转病毒的表观遗传学调控机制及其功能和非编码 RNA 调节机制及功能等。

内源性反转录病毒衍生的长非编码RNA的功能

胡序明, 崔恒宓*

(扬州大学动物科学与技术学院, 表观遗传学及表观基因组学研究所, 扬州 225009)

摘要: 长非编码 RNA (long non-coding RNA, lncRNA) 是一类转录本长度超过 200 nt、不编码蛋白质的 RNA 分子, 以 RNA 的形式参与多层次调控, 包括表观遗传学调控、转录调控以及转录后调控等。大量研究表明, 许多长非编码 RNA 分子衍生于数百万年前“入侵”人类基因组的内源性反转录病毒 (endogenous retrovirus, ERV) 序列。内源性反转录病毒是基因组重要成分, 约占基因组的 5%~8%, 多以“前病毒”形式存在, 功能很大程度上未知。就内源性反转录病毒衍生的 lncRNA 在天然免疫、抗病毒和肿瘤等方面的最新研究进展进行综述。

关键词: 长非编码 RNA; 内源性反转录病毒; 天然免疫; 表观遗传

中图分类号: Q522 **文献标志码:** A

Function of long non-coding RNA derived from endogenous retroviruses

HU Xu-Ming, CUI Heng-Mi*

(Institute of Epigenetics and Epigenomics, College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China)

Abstract: Long non-coding RNA (lncRNA) is a class of transcripts, which are non-protein coding RNA molecules longer than 200 nt. They have important regulatory functions in multi-levels, including epigenetic regulation, transcriptional and post-transcriptional regulations. A large number of studies indicate that many lncRNAs are derived from the endogenous retrovirus (ERV) sequences, which were thought to invade into the human genome millions of years ago. Endogenous retroviruses have become an important component of the genome, accounting for about 5%~8% of the genome. For these “proviral”, however, the function is largely unknown yet. In this paper, the latest research progress of lncRNA derived from ERVs in innate immunity, anti-virus and tumors are reviewed.

Key words: long non-coding RNA (lncRNA); endogenous retrovirus (ERV); innate immunity; epigenetics

收稿日期: 2016-03-26

基金项目: 国家自然科学基金重大研究计划(91540117); 国家自然科学基金面上项目(81372237)

*通信作者: E-mail: hmcui@yzu.edu.cn

人们已无从考证,为何英文名字 Virus 当初被译为病毒,其名使人将 Virus 视为又“病”又“毒”的坏东西。其实情况可能并非如此。就像细菌一样,虽然许多细菌与传染病有关,但也知道,细菌家族存在许多“益生菌”,人类离不开这些与人类有益的细菌。Virus 也是如此,虽然它们有些确实与传染病直接有关,但不少 Virus 并非又“病”又“毒”,而是与人类息息相关。正是对 Virus 的误解,人们对存在于生物基因组中的内源性病毒同样存在误解。内源性病毒分为先天性内源性病毒和后天性内源性病毒。先天性内源性病毒指长期进化过程遗传下来的内源性病毒序列,如人内源性反转录病毒和鸡内源性反转录病毒。由出生后病毒感染引起插入基因组,一般称为后天性内源性病毒,如乙肝病毒和 HIV 病毒感染导致的内源性插入性病毒。本文主要谈及先天性内源性反转录病毒。对先天性内源性反转录病毒的认识仍有争论。不少人认为内源性反转录病毒是有害的,但最新研究结果显示,内源性反转录病毒与早期胚胎发育有关,同时在胎盘中高表达,与胚胎干细胞全能性相关,且可能与细胞抗病毒有关。非编码 RNA 已成为当代生物学的研究热点,大量研究结果表明,许多长非编码 RNA 分子衍生于内源性反转录病毒序列的转录。本综述将着重讨论内源性反转录病毒衍生长非编码 RNA 及其作用和功能。

1 内源性反转录病毒

内源性反转录病毒 (endogenous retrovirus) 是基因组成分,约占基因组的 5%~8%,多以“前病毒”形式存在。内源性反转录病毒几乎在所有哺乳动物(如人、鼠、猫和羊)和脊椎动物(如鸡)基因组中存在。据称,内源性反转录病毒可能是远古反转录病毒感染 (retroviral infections) 宿主后的残留^[1]。过去对其功能不够了解,不少人曾将内源性病毒序列视为垃圾 DNA (junk DNA)^[2]。大部分内源性反转录病毒在进化过程中由于突变 (mutation)、缺失 (deletion) 等的积累已不表达或缺乏编码蛋白质的能力。最近有证据表明,一些内源性反转录病毒对胚胎发育、免疫应答具有作用,也与病毒感染以及肿瘤形成^[1,3-6]和胚胎干细胞的多能性有关^[5]。通常,内源性反转录病毒表达依赖于病毒长末端重复序列 (long terminal repeats, LTRs),而这些 LTRs 可以应答病毒和宿主的转录因子^[6-8]。LTR 具有转录内源性反转录病毒序列的启动子和增强子功能,此外,

还有转录子加尾作用^[7,9]。研究表明,内源性反转录病毒的表达不仅依赖于病毒长末端重复序列中的转录调节成分,还受到表观遗传学机制调控。

表观遗传学被定义为一种与 DNA 序列无关但能调节基因表达、影响表型的遗传现象。一方面,内源性反转录病毒基因受到严密控制,其表观遗传改变涉及癌症和自身免疫疾病。在诸多癌组织和细胞中都观察到内源性反转录病毒转录活性增强^[10],而去甲基化处理可以诱导内源性反转录病毒表达^[11]。研究证实, DNA 甲基化在沉默内源性反转录病毒中起着重要作用^[12]。在胚胎干细胞中,内源性反转录病毒沉默常常与组蛋白修饰有关^[13-14],被 H3K9me3 和 H4K20me3 所标记^[15]。并且,胚胎干细胞的全能性与某些内源性反转录病毒的活性紧密相关^[16]。另一方面,内源性反转录病毒在宿主细胞中通过重组可产生独立的 LTR 序列,这些 LTR 序列作为增强子、启动子等具有表观遗传调节基因表达的潜能^[9]。内源性反转录病毒的表观遗传调节可能与先天性免疫应答和病毒感染有关。研究表明,不同 Toll 样受体在配体刺激下可以激活内源性反转录病毒的表达;而 Nucleic Acid-Sensing Toll 样受体对控制内源性反转录病毒引起的病毒血症和诱导的肿瘤是必需的^[4]。机体受到外源性病毒,如反转录病毒 (HIV-1 和 HTLV-1)、疱疹病毒 (HSV-1 和 EBV) 感染也可以激活内源性反转录病毒的表达^[6,17-18]。这些病毒主要通过编码的病毒蛋白增加转录因子与内源性反转录病毒 LTR 序列上结合位点(如 NF- κ B 和 AP-1)的亲和力,从而反式激活内源性反转录病毒的表达^[6,19]。这表明先天性免疫系统和与炎症相关的转录因子能潜在调节内源性反转录病毒的表达。

有人认为,内源性反转录病毒可能与宿主共同进化了数百万年,经过正向选择以适应基因组的可塑性和进化,对防止相关病原及外源性反转录病毒感染具有保护作用。有趣的是,动物滋养层细胞分化和胎盘发育也有关,而胎盘滋养层细胞是正常机体内唯一类似肿瘤细胞的正常细胞。但是,关于内源性反转录病毒的表观遗传学的调控机制以及内源性反转录病毒在先天性免疫应答中与外源性反转录病毒之间的关系和相互作用,仍所知甚少。

目前已发现的反转录病毒种类繁多,但只有极少数反转录病毒在自然界同时以内源性和外源性两种形式存在,如禽白血病病毒 (avian leukosis virus, ALV)、鼠乳腺瘤病毒 (mouse mammary tumor viruses, MMTV)、猫白血病病毒 (feline leukemia virus, FLV)

和绵羊肺腺瘤病毒 (jaagsiekte sheep retrovirus, JSRV) 等^[20]。

2 长非编码RNA

长非编码 RNA (long non-coding RNA, lncRNA) 是非编码 RNA 的重要家族成员, 定义为转录本长度超过 200 nt 的 RNA 分子, 它们主要以 RNA 的形式参与多层次的调控, 包括表观遗传学调控、转录水平调控以及转录后 RNA 水平的调控。与小非编码 RNA, 如 miRNA 和 piRNA 一样, 它们是基因信息传递过程中基因表达时空性调控的有效调节因子。

lncRNA 最初被广泛认为是基因组垃圾 DNA (junk DNA) 转录的“噪音”(noise), 不具有生物学功能。但近年的研究证据表明, lncRNA 是细胞内基因信息传递过程中十分重要的调控因子, 参与许多重要生物学过程, 如 X 染色体沉默、基因组印记、组蛋白和 DNA 的修饰、转录抑制或激活等多种重要的调控。对 lncRNA 研究的极大兴趣还来自其与人类肿瘤的密切关系, 最近许多研究表明, lncRNA 与许多种肿瘤的多种致癌因素有关, 如肿瘤生长、转移、复发以及细胞凋亡^[21-26]。

2.1 lncRNA与表观遗传学调控

2.1.1 X染色体失活

在哺乳动物中, 雌性的一条 X 染色体完全沉默, 这种现象也称之为剂量补偿效应 (dosage compensation effect)^[27]。现已知道, lncRNA 直接与 X 染色体失活有关。X 染色体上一个 X 染色体失活中心 (X-inactivation center, Xic) 有一个 Xist 基因, 它能转录出 lncRNA, 直接控制 X 染色体失活^[28-29]。其转录的 Xist lncRNA 能引起 DNA 甲基化和相关组蛋白高甲基化和低乙酰化, 导致异染色质化和基因表达失活^[30]。有趣的是, Xist lncRNA 受反义 Tsix lncRNA 的负调节, Tsix lncRNA 能通过募集 DNMT3A DNA 甲基化酶到 Xist 启动子区, 抑制 Xist 基因的表达^[31-33]。

2.1.2 lncRNA与基因组印记(genomic imprinting)

现已发现, lncRNA 的转录也与一些印记 (imprinting) 基因调控有关, 而这些与印记调控有关的 lncRNA 都是印记基因的反义 lncRNA。一个例子是与 BWS 综合征相关印记区的 KCNQOT1 lncRNA, 该基因位于 KCNQ1 基因第 10 内含子内, 转录子为 KCNQ1 的反义 lncRNA, 直接控制 KCNQ1 的印记。有趣的是, 正义 KCNQ1 仅从母源等位基因表达, 反义 KCNQ1 lncRNA 仅从父源等位基因

表达, 且反义 KCNQOT1 lncRNA 通过影响 KCNQ1 的印记控制区 (imprinting control region, ICR) 的 DNA 甲基化来控制 KCNQ1 基因印记^[34-37]。

2.2 lncRNA与转录水平调控

lncRNA 可在转录水平调控基因表达, 其调控方式包括: (1) lncRNA 的转录可干扰附近基因的表达, 如酵母 SER3 基因会受到上游 lncRNA SRG1 转录的干扰^[38]; (2) lncRNA 可通过封阻启动子区域来干扰基因的表达, 如 DHFR 上游的一个 lncRNA 能与 DHFR 启动子形成 RNA-DNA 螺旋结构, 从而抑制转录因子 TFIID 的结合, 抑制 DHFR 基因表达^[39]; (3) lncRNA 可与 RNA 结合蛋白作用, 并将其募集到基因启动子区从而调控基因表达, 如 CCND1 启动子上游一个 lncRNA 可调节 RNA 结合蛋白 TLS 的活性, 进而调控 CCND1 基因表达^[40]; (4) lncRNA 能调节转录因子活性, 如 lncRNA Evf2 能与转录因子 Dlx2 形成转录复合物从而激活 Dex6 的表达^[41-42]; (5) Alu RNA 能通过抑制 RNA 聚合酶 II 来实现广泛的基因抑制^[43-44]。

2.3 lncRNA与转录后调控

lncRNA 也能在转录后水平通过与 mRNA 形成双链来调节 mRNA 水平, 如 p53 反义 RNA (Wrap53) 的 3' 端可与 p53 mRNA 3' 端形成双链, 从而增加了 p53 mRNA 的稳定性^[45]。显然, lncRNA 是长期进化过程中形成的基因调控网络中不可或缺的调节因子。它们既可在胞浆, 又可在胞核中通过不同的作用机制, 行使调节功能。在胞浆中, lncRNA 主要行使转录后水平的调节作用, 这包括通过改变编码基因 mRNA 的稳定性, 改变翻译效率来调节基因表达水平; 也能通过 miRNA 通路, 如作为 miRNA 前体或作为内源性竞争性 RNA (ceRNA) 来调节编码基因 mRNA 的水平^[46]。此外, mRNA 水平的调节也可通过与蛋白质编码基因的转录本形成互补双链, 在 Dicer 酶作用下产生内源性 siRNA。lncRNA 也可与特定蛋白结合, 改变蛋白在胞质中定位, 或调节相应蛋白的活性^[47]。

lncRNA 对表观遗传学调控以及转录水平的调控均发生在胞核中。lncRNA 可募集染色质重构复合体, 如 PCR2 复合体到特定位点, 引起 DNA 甲基化或组蛋白修饰变化, 从而介导相关基因的沉默。例如, HOTAIR、p15AS、Xist、AIR、KCNQLOT1 等 lncRNA 均可募集染色质重构复合体, 通过其中的组蛋白甲基转移酶, 如 Ezh2 或 G9a 实现表观遗传学沉默^[48]。不久前, Chu 等^[49]报道了用 RNA 纯

化染色质分离 (chromatin isolation by RNA purification, ChIRP) 技术研究 lncRNA 与其他大分子的结合和相互作用。显然, 此技术对研究 lncRNA 的作用机制具有重要贡献^[22]。

3 源自内源性反转录病毒的lncRNA

至今已有 1 万多个 lncRNA 被报道^[50]。与编码 RNA 相比, lncRNA 表达水平偏低且具有组织特异性^[51-53]。按照 lncRNA 在基因组上相对于蛋白质编码基因的位置、转录方向和功能, 可将 lncRNA 分为独立 lncRNA (stand-alone lncRNA)、非编码反义 RNA、内含子 RNA (intronic RNA)、基因间 RNA (intergenic RNA)、启动子相关 RNA (promoter-associated RNA)、增强子 RNA (enhancer RNA) 和假基因 RNA (pseudogene RNA) 等^[54]。有研究表明, 基因组中内源性病毒也是 lncRNA 的一个重要来源^[55-56]。近年来, 人们发现内源性反转录病毒可产生或者诱产生长链非编码 RNA^[5,55,57-58]。大约 10% 由人内源性反转录病毒 H 亚科 (human endogenous retrovirus subfamily H, HERVH) 衍生的转录物变为 lncRNA^[55,57]。据报道, HERVH 作为一种核内长非编码 RNA, 可通过与 OCT4 和共激活因子作用以及激活长非编码 RNA 影响胚胎干细胞发育^[5,57]。内源性反转录病毒相关长链非编码 RNA (endogenous retrovirus-associated long non-coding RNA, ERV-lncRNAs) 则在肿瘤中被激活^[59]。在鸡长非编码 RNA 数据库 (domestic-animal lncRNA database, ALDB) 中, 发现 2 条来自鸡 1 号染色体负链上由内源性反转录病毒 ALVE1 衍生的长链非编码 RNA ALDBGALT0000000876 和 ALDB-GALT-0000000877, 其长度分别为 1 404 nt 和 1 655 nt^[60]。

另外, 源自内源性反转录病毒序列的转座子 (transposable elements, TEs) 则可能与脊椎动物 lncRNA 的起源、多样化及其调节功能密切相关。据报道, TE 普遍存在于脊椎动物的 lncRNA 序列中^[61]。这些插入的 TE 可能是 lncRNA 的关键功能区域, 如参与 RNA、DNA 和蛋白质结合^[62]。在进化过程中, 一部分 TE 插入是 lncRNAs 多样化的重要来源, 而其他一些 TE 插入可能导致全新 lncRNAs 的出现。究竟先有 lncRNA, 还是先有 TE? 目前, 暂无确切答案。

所以认为, 内源性反转录病毒是宿主长非编码 RNA 的重要来源之一, 不仅可以在多种层面上 (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, 而且也具备抵抗外源性病毒的分子基

础。在内源性与外源性病毒相互作用研究中, 人们发现内源性反转录病毒可能会阻断外源性病毒的感染和复制^[63-64]。尽管目前还没有直接证据显示内源性反转录病毒产生的 lncRNA 可以抵抗外源性病毒感染, 但这一推测在外源性病毒研究中得到证实。外源性反转录病毒 HIV 能够编码反义 lncRNA, 并且通过表观遗传学机制调控 HIV 病毒转录^[65]。这增加了人们对内源性反转录病毒产生的 lncRNA 的功能及其作用机制的关注。内源性病毒产生的 lncRNA 功能研究将是一个全新的领域, 很可能帮助揭示内源性病毒抗病机制的神秘面纱。

4 内源性反转录病毒衍生的长非编码RNA的功能

4.1 天然免疫

近年来, lncRNA 在天然免疫中的作用也逐渐被认识^[66-67]。起初, 人们发现脂多糖 (lipopolysaccharides, LPS) 刺激小鼠骨髓树突状细胞后引起 lncRNA 异常表达, 随后证实这些 lncRNA 与小鼠肺部的病毒感染、单核细胞和巨噬细胞的激活密切相关^[68]。在人和鼠巨噬细胞中, lincRNA-Cox2 和 THRIL 对调节 TLR2 信号起着关键作用。最新研究表明, TLR3 配体刺激导致大量 lncRNA 异常表达, 暗示 lncRNA 可能作为 TLR3 信号转导的一部分参与天然免疫调节^[69]。长非编码 RNA linc-MAF-4 可以调节 T 淋巴细胞分化^[70], 而 lnc-DC 通过转录因子 STAT3 控制人类树突状细胞分化^[71]。NEAT1、NeST 和 THRIL 这三种 lncRNA 分别通过调控 IL-8、IFN- γ 和 TNF- α 的转录来调节细胞天然免疫反应^[72-74]。一种新发现的长非编码 RNA NRAV 可通过影响干扰素刺激关键基因, 如 IFITM3 和 MXA 的起始转录参与宿主抗病毒天然免疫反应^[75]。因此, lncRNA 可作为 Toll 样受体信号和天然免疫的调节因子^[76]。

那么, 源于内源性反转录病毒的 lncRNA 是否在天然免疫中发挥作用呢? 研究表明, 天然免疫相关的转录因子能反式激活内源性反转录病毒表达^[6]; 相反, 内源性反转录病毒也可能通过病原模式识别受体 (如 Toll 样受体) 激活天然免疫^[4,77]。内源性反转录病毒的激活表达可能会产生一些具有病毒特征的 RNA (如 dsRNA、lncRNAs) 或蛋白质, 因而可以像外源性病毒的病原相关分子模式一样被宿主天然免疫系统识别^[78]。已有证据表明, 人内源性反转录病毒 HERV 可以与 Toll 样受体相互作用, 并且其核酸产物可以激活细胞病原模式识别受体。

另外, 有研究表明, 去甲基化试剂 AZA 处理可以激活内源性反转录病毒的双向转录, 从而形成 dsRNA 二级结构诱导干扰素应答反应^[79]。因此, 完全有理由相信, 内源性反转录病毒衍生的 lncRNAs 是宿主天然免疫系统的组成部分, 并参与调节宿主免疫应答反应。

4.2 抗病毒作用

lncRNAs 作为一类重要的 RNA 调节分子, 已被证明参与抗病毒天然免疫反应^[80]。高通量测序分析揭示, 病毒感染引起一系列的宿主 lncRNAs 的异常表达。功能获得或缺失实验证实, 一些 lncRNAs (如 NeST、Lethe、lncRNA-CMPK2、VIN 和 NRON) 被病毒所控制以利于病毒感染, 而其他一些 lncRNAs (如 lincRNACox2、NRAV、NEAT1、7SL 和 lnc-DC) 在抗病毒反应中具有非常关键的作用^[80]。并且, 在病毒感染过程中, 宿主 lncRNAs 的表达可以被病毒编码的组分或者病毒诱导细胞因子、干扰素和趋化因子等所调节。这些差异表达的 lncRNAs 可能正向或负向调节抗病毒反应的诸多关键环节。关键的天然免疫分子的表达变化显著影响了宿主抗病毒反应, 因此影响病毒感染和复制。

值得注意的是, 一些在天然免疫中具有负调控作用的 lncRNAs 被病毒所控制以抑制抗病毒反应, 具有正调控作用的 lncRNAs 则被病毒所抑制。例如, NRAV 是一种新发现的长非编码 RNA, 主要通过影响干扰素刺激关键基因, 如 IFITM3 和 MXA 的起始转录参与宿主抗病毒天然免疫反应^[75]。一些病毒, 如甲型流感病毒 (influenza A virus)、简单疱疹病毒 (herpes simplex virus)、番鸭呼肠孤病毒 (muscovy duck reovirus) 和仙台病毒 (sendai virus) 显著下调 NRAV 表达。NRAV 的下调导致一些干扰素刺激基因, 如 MxA、IFITM3、IFIT2、IFIT3 和 OASL 被诱导表达, 从而显著破坏病毒的复制^[75]。另外, 干扰素诱导的 lncRNA-CMPK2 被发现可作为一系列干扰素刺激基因, 如 ISG15、CXCL10、IFIT3、IFITM1、Viperin、CMPK2 等的负调控因子。lncRNA-CMPK2 在丙肝病毒 (HCV) 感染的肝脏组织中显著上调, 这可能有利于 HCV 复制^[81]。

在病毒感染中, 宿主 lncRNAs 可能调节抗病毒反应的不同过程, 包括转录因子的活性和迁移、干扰素和细胞因子的产生、干扰素刺激基因的转录及其免疫细胞的发育等等。尽管目前关于内源性反转录病毒衍生的 lncRNAs 在抗病毒反应中的作用知之甚少, 但基于本实验室的研究结果可以肯定内源

性反转录病毒衍生的 lncRNAs 具有抗病毒功能。另外, 这一观点也在外源性病毒研究中得到证实。外源性反转录病毒 HIV 能够编码反义 lncRNA, 并且通过表观遗传学机制调控 HIV 病毒转录^[65]。因此, 可以想象, 内源性反转录病毒衍生的 lncRNAs 可以通过调节抗病毒天然免疫反应或反义 lncRNAs, 抵抗外源性病毒的感染和复制。

4.3 抗癌作用

对 lncRNA 研究的极大兴趣还来自其与人类肿瘤的密切关系, 最近许多研究结果表明 lncRNA 与肿瘤的多种致癌因素有关, 如肿瘤生长、转移等^[82-83]。研究表明, 内源性反转录病毒及其相关的 lncRNA 在正常组织或细胞中保持沉默, 而在肿瘤组织或细胞中被异常激活^[59]。令人惊讶的是, 外源性病毒核苷酸序列和 lncRNA 在正常细胞中均保持沉默, 而在肿瘤细胞中被转录激活。在这种“特殊”肿瘤环境中, 被激活的内源性反转录病毒可能在抗肿瘤免疫中具有非常关键的调节作用^[84]。最新研究指出, 抗癌药物去甲基化试剂 AZA 通过内源性反转录病毒激活了肿瘤细胞的天然免疫反应^[79], 而激活天然免疫可以提高机体抗癌或抗肿瘤免疫能力。内源性反转录病毒相关的 lncRNAs 在黑色素瘤中则被激活, 所以内源性反转录病毒衍生的 lncRNAs 可能具有抗癌作用或者可作为肿瘤标志物, 这解释了内源性反转录病毒及其衍生的 lncRNA 在肿瘤中为什么异常激活。

4.4 其他

内源性反转录病毒衍生的长非编码 RNA 还可能与胚胎干细胞发育有关。据报道, HERVH 作为一种核内长非编码 RNA, 可通过与 OCT4 和共激活因子作用影响胚胎干细胞发育^[5]。科学家已经发现, 数百万年前“入侵”人类基因组的病毒已经改变了人类胚胎干细胞 (ES 细胞) 中的基因开启与关闭方式^[85]。通过运用新的测序技术对人类和小鼠胚胎干细胞 (ES 细胞) 中三种调节蛋白质 (OCT4、NANOG 与 CTCF) 的染色体组定位 (基因组定位) 的研究发现, 数百万年前自行插入人类基因组的特定类型病毒已经戏剧性地改变了人类干细胞基因调控网络^[85]。在胚胎干细胞中, 内源性反转录病毒的转录活性与表观遗传修饰, 如 DNA 甲基化和组蛋白修饰有关^[13-14], 常被 H3K9me3 和 H4K20me3 所标记^[15]。并且, 胚胎干细胞的潜能与内源性反转录病毒的活性紧密相关^[16]。内源性反转录病毒衍生的长非编码 RNA 可能作为关键的调节因子参与胚胎

干细胞的发育, 并赋予了其特殊的功能。

5 结语与展望

长非编码 RNA 已成为当代生物学的研究热点, 并且越来越多的研究表明, 许多长非编码 RNA 分子衍生于内源性反转录病毒序列的转录。数百万年前“入侵”人类基因组的病毒可能极大地参与了细胞基因调控网络的调节, 如调节细胞发育、天然免疫以及肿瘤发生发展等。研究这些内源性反转录病毒的 lncRNA 的功能和作用将是一个全新的领域, 不仅可以帮助揭示内源性病毒的真正功能和作用, 也有助于理解内源性反转录病毒参与的复杂的人类基因组表观遗传学调控机制。

[参 考 文 献]

- [1] Weiss RA. The discovery of endogenous retroviruses. *Retrovirology*, 2006, 3: 67
- [2] Villarreal LP. Viral ancestors of antiviral systems. *Viruses*, 2011, 3: 1933-58
- [3] Kurth R, Bannert N. Beneficial and detrimental effects of human endogenous retroviruses. *Int J Cancer*, 2010, 126: 306-14
- [4] Yu P, Lubben W, Slomka H, et al. Nucleic acid-sensing Toll-like receptors are essential for the control of endogenous retrovirus viremia and ERV-induced tumors. *Immunity*, 2012, 37: 867-79
- [5] Lu X, Sachs F, Ramsay L, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*, 2014, 21: 423-5
- [6] Manghera M, Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology*, 2013, 10: 16
- [7] Khodosevich K, Lebedev Y, Sverdlov E. Endogenous retroviruses and human evolution. *Comp Funct Genomics*, 2002, 3: 494-8
- [8] Moyes D, Griffiths DJ, Venables PJ. Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. *Trends Genet*, 2007, 23: 326-33
- [9] Buzdin A. Human-specific endogenous retroviruses. *ScientificWorldJournal*, 2007, 7: 1848-68
- [10] Ruprecht K, Mayer J, Sauter M, et al. Endogenous retroviruses and cancer. *Cell Mol Life Sci*, 2008, 65: 3366-82
- [11] Leroy V, Kihara M, Baudino L, et al. Sgp3 and TLR7 stimulation differentially alter the expression profile of modified polytropic retroviruses implicated in murine systemic lupus. *J Autoimmun*, 2012, 38: 361-8
- [12] Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 2004, 431: 96-9
- [13] Matsui T, Leung D, Miyashita H, et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, 2010, 464: 927-31
- [14] Rowe HM, Jakobsson J, Mesnard D, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 2010, 463: 237-40
- [15] Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 2007, 448: 553-60
- [16] Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, 2012, 487: 57-63
- [17] Assinger A, Yaiw KC, Gottesdorfer I, et al. Human cytomegalovirus (HCMV) induces human endogenous retrovirus (HERV) transcription. *Retrovirology*, 2013, 10: 132
- [18] Bhardwaj N, Maldarelli F, Mellors J, et al. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses in vivo but not to increased virion production. *J Virol*, 2014, 88: 11108-20
- [19] Gonzalez-Hernandez MJ, Cavalcoli JD, Sartor MA, et al. Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 Tat protein. *J Virol*, 2014, 88: 8924-35
- [20] Gifford R, Tristem M. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes*, 2003, 26: 291-315
- [21] Hu X, Feng Y, Zhang D, et al. A functional genomic approach identifies *FALI* as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell*, 2014, 26: 344-57
- [22] Barnhill LM, Williams RT, Cohen O, et al. High expression of CAI2, a 9p21-embedded long noncoding RNA, contributes to advanced-stage neuroblastoma. *Cancer Res*, 2014, 74: 3753-63
- [23] Yuan JH, Yang F, Wang F, et al. A long noncoding RNA activated by TGF- β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*, 2014, 25: 666-81
- [24] Yang L, Lin C, Jin C, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*, 2013, 500: 598-602
- [25] Du Z, Fei T, Verhaak RG, et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol*, 2013, 20: 908-13
- [26] Yang F, Huo XS, Yuan SX, et al. Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell*, 2013, 49: 1083-96
- [27] Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*, 1961, 190: 372-3
- [28] Rastan S. X chromosome inactivation and the *Xist* gene. *Curr Opin Genet Dev*, 1994, 4: 292-7
- [29] Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat Genet*, 2002, 30: 167-74
- [30] Zhao J, Sun BK, Erwin JA, et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.

- Science, 2008, 322: 750-6
- [31] Lee JT, Davidow LS, Warshawsky D. *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nat Genet*, 1999, 21: 400-4
- [32] Bacher CP, Guggiari M, Brors B, et al. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol*, 2006, 8: 293-9
- [33] Xu N, Tsai CL, Lee JT. Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, 2006, 311: 1149-52
- [34] Fitzpatrick GV, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of *KvDMRI*. *Nat Genet*, 2002, 32: 426-31
- [35] Thakur N, Tiwari VK, Thomassin H, et al. An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol Cell Biol*, 2004, 24: 7855-62
- [36] Smilnich NJ, Day CD, Fitzpatrick GV, et al. A maternally methylated CpG island in *KvLQTI* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci USA*, 1999, 96: 8064-9
- [37] Mitsuya K, Meguro M, Lee MP, et al. *LIT1*, an imprinted antisense RNA in the human *KvLQTI* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum Mol Genet*, 1999, 8: 1209-17
- [38] Thebault P, Boutin G, Bhat W, et al. Transcription regulation by the noncoding RNA *SRGI* requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol Cell Biol*, 2011, 31: 1288-300
- [39] Shi X, Sun M, Liu H, et al. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett*, 2013, 339: 159-66
- [40] Song X, Wang X, Arai S, et al. Promoter-associated noncoding RNA from the *CCND1* promoter. *Methods Mol Biol*, 2012, 809: 609-22
- [41] Berghoff EG, Clark MF, Chen S, et al. *Evf2 (Dlx6as)* lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development*, 2013, 140: 4407-16
- [42] Bond AM, Vangompel MJ, Sametsky EA, et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci*, 2009, 12: 1020-7
- [43] Yakovchuk P, Goodrich JA, Kugel JF. B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci USA*, 2009, 106: 5569-74
- [44] Mariner PD, Walters RD, Espinoza CA, et al. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell*, 2008, 29: 499-509
- [45] Mahmoudi S, Henriksson S, Corcoran M, et al. *Wrap53*, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell*, 2009, 33: 462-71
- [46] Zhang K, Shi ZM, Chang YN, et al. The ways of action of long non-coding RNAs in cytoplasm and nucleus. *Gene*, 2014, 547: 1-9
- [47] Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 2009, 23: 1494-504
- [48] Redrup L, Branco MR, Perdeaux ER, et al. The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development*, 2009, 136: 525-30
- [49] Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp*, 2012, pii: 3912
- [50] Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet*, 2014, 30: 348-55
- [51] Marques AC, Ponting CP. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol*, 2009, 10: R124
- [52] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*, 2012, 22: 1775-89
- [53] Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*, 2011, 25: 1915-27
- [54] Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*, 2013, 193: 651-69
- [55] Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol*, 2012, 13: R107
- [56] Faulkner GJ, Kimura Y, Daub CO, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet*, 2009, 41: 563-71
- [57] Wang J, Xie G, Singh M, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, 2014, 516: 405-9
- [58] Fasching L, Kapopoulou A, Sachdeva R, et al. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep*, 2015, 10: 20-8
- [59] Gibb EA, Warren RL, Wilson GW, et al. Activation of an endogenous retrovirus-associated long non-coding RNA in human adenocarcinoma. *Genome Med*, 2015, 7: 22
- [60] Li A, Zhang J, Zhou Z, et al. ALDB: a domestic-animal long noncoding RNA database. *PLoS One*, 2015, 10: e0124003
- [61] Kapusta A, Kronenberg Z, Lynch VJ, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*, 2013, 9: e1003470
- [62] Johnson R, Guigo R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA*, 2014, 20: 959-76
- [63] Tandon R, Cattori V, Pepin AC, et al. Association between endogenous feline leukemia virus loads and exogenous feline leukemia virus infection in domestic cats. *Virus Res*, 2008, 135: 136-43

- [64] Murcia PR, Arnaud F, Palmarini M. The transdominant endogenous retrovirus enJS56A1 associates with and blocks intracellular trafficking of Jaagsiekte sheep retrovirus Gag. *J Virol*, 2007, 81: 1762-72
- [65] Saayman S, Ackley A, Turner AM, et al. An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol Ther*, 2014, 22: 1164-75
- [66] Aune TM, Spurlock CF 3rd. Long non-coding RNAs in innate and adaptive immunity. *Virus Res*, 2016, 212: 146-60
- [67] Satpathy AT, Chang HY. Long noncoding RNA in hematopoiesis and immunity. *Immunity*, 2015, 42: 792-804
- [68] Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 2009, 458: 223-7
- [69] Wang S, Li X, Zhao RC. Transcriptome analysis of long noncoding RNAs in Toll-like receptor 3-activated mesenchymal stem cells. *Stem Cells Int*, 2016, 2016: 6205485
- [70] Ranzani V, Rossetti G, Panzeri I, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol*, 2015, 16: 318-25
- [71] Wang P, Xue Y, Han Y, et al. The STAT3-binding long noncoding RNA linc-DC controls human dendritic cell differentiation. *Science*, 2014, 344: 310-3
- [72] Imamura K, Imamachi N, Akizuki G, et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol Cell*, 2014, 53: 393-406
- [73] Gomez JA, Wapinski OL, Yang YW, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- γ locus. *Cell*, 2013, 152: 743-54
- [74] Li Z, Chao TC, Chang KY, et al. The long noncoding RNA *THRIL* regulates TNF α expression through its interaction with hnRNPL. *Proc Natl Acad Sci USA*, 2014, 111: 1002-7
- [75] Ouyang J, Zhu X, Chen Y, et al. NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription. *Cell Host Microbe*, 2014, 16: 616-26
- [76] Murphy MB, Medvedev AE. Long noncoding RNAs as regulators of Toll-like receptor signaling and innate immunity. *J Leukoc Biol*, 2016, 99: 839-50
- [77] Blasius AL, Beutler B. Intracellular toll-like receptors. *Immunity*, 2010, 32: 305-15
- [78] Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. *J Gen Virol*, 2015, 96: 1207-18
- [79] Chiappinelli KB, Strissel PL, Desrichard A, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell*, 2015, 162: 974-86
- [80] Ouyang J, Hu J, Chen JL. lncRNAs regulate the innate immune response to viral infection. *Wiley Interdiscip Rev RNA*, 2016, 7: 129-43
- [81] Kambara H, Niazi F, Kostadinova L, et al. Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res*, 2014, 42: 10668-80
- [82] Trimarchi T, Bilal E, Ntziachristos P, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell*, 2014, 158: 593-606
- [83] Rupaimoole R, Lee J, Haemmerle M, et al. Long noncoding RNA ceruloplasmin promotes cancer growth by altering glycolysis. *Cell Rep*, 2015, 13: 2395-402
- [84] Tanne A, Muniz LR, Puzio-Kuter A, et al. Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells. *Proc Natl Acad Sci USA*, 2015, 112: 15154-9
- [85] Kunarso G, Chia NY, Jeyakani J, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*, 2010, 42: 631-4