

DOI: 10.13376/j.cbls/2016079

文章编号: 1004-0374(2016)05-0615-07



汪香婷, 中国科学技术大学生命科学学院研究员, 博士生导师。研究组正式组建于2014年9月, 主要通过生化和分子生物学、细胞生物学以及基因编辑等技术, 对长链非编码RNA在DNA损伤修复以及与DNA损伤修复异常相关的重大人类疾病中的功能和作用分子机制进行离体和在体动物研究。我们的研究从临床迫切需要解决的问题入手, 希望能够发掘一些具有很强特异性的长链非编码RNA疾病标记物, 并将成果运用于肿瘤等严重危害人类健康的疾病的诊断和治疗。

## 长链非编码RNA在体动物研究进展

申涛, 吴向琴, 郭玉珠, 汪香婷\*

(中国科学技术大学生命科学学院, 中科院分子细胞科学卓越创新中心, 中科院脑功能与脑疾病重点实验室, 合肥 230026)

**摘要:** 从一开始被怀疑为基因组的噪音到目前成为生物医学研究领域的宠儿, 主要基于细胞水平的长链非编码RNA (long noncoding RNA, lncRNA) 研究在过去的10年左右时间里得到了迅猛发展。随着近几年在体动物研究的发展, lncRNA领域正在由其新生阶段进入下一个重要发展阶段。然而, 由于lncRNA的特殊性, 使得lncRNA在体动物研究难度依然很大, 对实验策略, 甚至目的基因的选择都有很强的技巧。现就lncRNA在体动物研究现状和其中一些重要问题进行综述。

**关键词:** 长链非编码RNA; 在体动物研究; Hotair; Malat1/Neat2; Fendrr; CRISPR-Cas9

**中图分类号:** Q527 **文献标志码:** A

### Recent progresses of long noncoding RNA *in vivo* study

SHEN Tao, WU Xiang-Qin, GUO Yu-Zhu, WANG Xiang-Ting\*

(CAS Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences Key Laboratory of Brain Function and Disease, School of Life Sciences, University of Science and Technology of China, Hefei 230026, China)

**Abstract:** From initially being considered as some genome noise to currently as one of the most extensively investigated topics in biomedical science, the long noncoding RNA (lncRNA) field has obtained a rapid growth during the past ten years or so, mainly based on studies in cultured cells. With the recent progresses on lncRNA *in vivo* investigations, the lncRNA field is stepping into a new era from its infancy. However, there is still a big challenge to the lncRNA *in vivo* study field, due to the distinction between lncRNA and the protein-encoded genes. Well consideration and appropriate approach are needed when choosing a modification strategy, or if possible, when choosing a targeted gene. We will discuss the research progresses of lncRNA *in vivo* study and several important issues in this review.

**Key words:** long noncoding RNA; *in vivo* study; Hotair; Malat1/Neat2; Fendrr; CRISPR-Cas9

收稿日期: 2015-02-18

基金项目: 国家自然科学基金项目(31471226); 基金委重大研究计划项目(91440108); 教育部基本科研业务费专项资金青年创新基金(WK2070000044); 教育部基本科研业务费专项资金创新团队培育基金(WK2070000034)

\*通信作者: E-mail: wangxt11@ustc.edu.cn

长链非编码 RNA (long noncoding RNA, lncRNA) 指长度大于 200 个核苷酸的非编码 RNA, 在多种细胞生理和病理过程中的重要作用已被广泛接受<sup>[1-5]</sup>。尽管 lncRNA 研究是近年来的新兴领域, 对 lncRNA 的研究却可以追溯到 20 世纪 70 年代的 7SK<sup>[6]</sup> 以及 90 年代初两个基因组印记基因 (imprinting gene) *H19*<sup>[7]</sup> 和 *Xist*<sup>[8]</sup> 的发现。继 *H19* 和 *Xist* 之后又陆续地发现了一些在 X 染色体失活和剂量补偿中发挥作用的印记 lncRNAs, 如 *Air*<sup>[9]</sup> 和 *Kcnqlot1*<sup>[10]</sup> 等。而非基因组印记相关的 lncRNAs 则被发现在生物进化、干细胞维持、胚胎发育、细胞分化、凋亡、代谢、信号转导、感染以及免疫应答等诸多方面发挥重要调控作用。与蛋白质编码基因相比, lncRNA 具有四大特征: 其一, 数量庞大, 具有复杂的种类和作用机制; 其二, 显著的时空和组织特异性, 常表达于细胞核中; 其三, 在进化过程中较低的保守性; 其四, 对信号刺激和环境变化的高敏感性和可塑性。目前对 lncRNA 功能的认识主要源于以培养的细胞为实验材料的研究工作。由于 lncRNA 的特殊特征, 尤其是其较高的时空表达特性, 使 lncRNA 在体动物研究有着比蛋白质编码基因更为迫切的需求。

目前已知的 lncRNA 数量达上万种之多, 有功能报道的亦有数百种, 但是 lncRNA 在体动物研究仅有 30 余例 (表 1 和 2)<sup>[11-38]</sup>, 而且在此计算中由不同研究组针对同一 lncRNA 的研究报道被累加计算。一方面, 相对于较为成熟的蛋白质编码基因研究领域, lncRNA 研究刚刚起步; 另一方面则是由于对

lncRNA 的认识不足和技术手段局限性等难题, 使 lncRNA 在体动物研究遇到很多障碍。然而, 令人欣喜的是, 经过不懈的探索 and 知识积累, lncRNA 在体动物研究在最近几年有了较大突破。在如何合理选择目的基因和设计研究策略方面都有了较为系统的认识。此外, 近几年新发展起来的技术手段, 如 CRISPR-Cas9 基因编辑技术<sup>[39-40]</sup> 亦有望成为促进 lncRNA 在体研究发展的强有力工具。因此, 从整体动物水平上探索 lncRNA 功能, 推动着 lncRNA 领域由其新生阶段进入其下一个重要发展阶段。本文将首先回顾 lncRNA 在体动物研究发展的历程, 以几个有代表性的基因沉默小鼠为例讨论该领域的难点, 最后对解决方案进行总结和讨论。

## 1 lncRNA在体动物研究发展简程

### 1.1 基因组印记基因占主导地位的起始阶段

从 1990 年普林斯顿大学 Tilghman 研究组发现 *H19*<sup>[7]</sup> 到 2003 年 *Malat1*<sup>[41]</sup> 被报道前的 10 余年间, 基因组印记领域在 lncRNA 研究中独占鳌头。这期间, 共有 9 例与基因组印记相关的 lncRNA 基因沉默小鼠和 1 例果蝇相关研究问世 (表 1)。采用传统的基因敲除策略, *Xist*<sup>[14-15]</sup>、*Kcnqlot1*<sup>[19]</sup>、*Air*<sup>[21]</sup> 和 *Tsix*<sup>[22-23]</sup> 基因沉默小鼠均获得了与预期基本相符的表型。基因组印记基因研究对 lncRNA 领域发展做出了十分重要的贡献。

2003—2009 年基本上为 lncRNA 在体动物研究的空档期。这一阶段初期, 在培养的细胞上, 一些

表1 基因组印记lncRNA失活的案例

LncRNA	动物模型	修饰策略	观察到的表型	参考文献
<i>H19</i>	鼠	13 kb (3 kb基因+10 kb上游)置换	胎盘和胎儿的重量增加	[11]
<i>H19</i>	鼠	增强子敲除	微弱的系统性胎儿生长阻滞	[12]
<i>H19</i>	鼠	3 kb转录元件置换	胎盘和胎儿轻微的重量增加	[13]
<i>Xist</i>	鼠	15 kb置换	从母系继承来的父本等位基因导致的胚胎期致死	[14]
<i>Xist</i>	鼠	全基因敲除	正常的X染色体失活	[15]
<i>Xist</i>	鼠	外显子1到外显子5倒置	从父本继承来的等位基因导致的胚胎期致死	[16]
<i>Xist</i>	鼠	全基因敲除	血癌	[17]
<i>roX</i>	黑腹果蝇	敲除roX 1/2	正常	[18]
<i>Kcnqlot1</i>	鼠	敲除启动子	从父本继承来的等位基因导致的生长缺陷	[19]
<i>Kcnqlot1</i>	鼠	提前转录终止	生长缺陷	[20]
<i>Air</i>	鼠	提前转录终止	从父本继承来的等位基因导致的生长缺陷	[21]
<i>Tsix</i>	鼠	敲除启动子	母系基因传递受损、生长阻滞	[22]
<i>Tsix</i>	鼠	提前转录终止	胚胎期致死	[23]
<i>Tsix</i>	鼠	2 kb的启动子和外显子1敲除	雄鼠发生睾丸变小, 恐惧本能减弱的突变	[24]
<i>Jpx</i>	鼠	启动子和外显子敲除	胚胎期致死	[25]
<i>Rian</i>	鼠	CRISPR/Cas9	组织特异性的顺式调控邻近基因	[26]

表2 非基因组印记lncRNA失活的案例

LncRNA	动物模型	修饰策略	观察到的表型	参考文献
<i>BCI</i>	鼠	启动子和外显子置换	对听觉刺激反应缺陷	[27]
<i>Eyf2</i>	鼠	提前转录终止	正常	[28]
<i>Neat1</i>	鼠	启动子及上游基因敲除	正常	[29]
<i>Malat1/Neat2</i>	鼠	提前转录终止	正常	[30]
<i>Malat1/Neat2</i>	鼠	启动子及上游基因敲除	正常	[31]
<i>Malat1/Neat2</i>	鼠	全基因敲除	正常	[32]
<i>Hotair</i>	鼠	HoxC基因簇敲除	正常	[33]
<i>Hotair</i>	鼠	全基因敲除	同源异形转化、骨骼畸形、基因去抑制	[34]
<i>Hog and Tog</i>	鼠	基因倒置	盲肠发育阶段HoxD表达缺失	[35]
<i>Fendrr</i>	鼠	提前转录终止	胚胎期致死, 心脏、体壁发育缺陷	[36]
<i>Fendrr</i>	鼠	基因座位点置换	围产期致死、多种器官缺陷, 包括肺部化脓和血管化	[37]
<i>Peril</i>	鼠	基因座位点置换	围产期致死	[37]
<i>Mdgt</i>	鼠	基因座位点置换	围产期致死、生长减缓	[37]
<i>linc-Brn1b</i>	鼠	基因座位点置换	大脑皮层发育异常	[37]
<i>Spasm</i>	鼠	基因座位点置换	行为学紊乱、颤动	[37]
<i>linc-p21</i>	鼠	基因座位点置换	行为学紊乱	[37]
<i>NeST</i>	鼠	自然多态性或转基因	对特异的细菌性疾病免疫增强	[38]

当时还未以 lncRNA 命名的非编码 RNA 的作用机制被报道, 它们大多数在基因转录调控水平发挥作用<sup>[1]</sup>。该阶段末期为 lncRNA 领域形成时期, 大量以 lncRNA 命名的非编码 RNA 被发现<sup>[42-43]</sup>。

## 1.2 非基因组印记lncRNA基因发展阶段

2009年后, 随着 *BCI*<sup>[27]</sup> 和 *Eyf2*<sup>[28]</sup> 基因沉默小鼠的报道, 非基因组印记相关 lncRNA 在体研究开始呈现连续性发展(表2)。在此期间, 一些 lncRNA 领域明星基因, 如 *Neat1*<sup>[34]</sup>、*Malat1/Neat2*<sup>[30-32]</sup> 和 *Hotair*<sup>[33-34]</sup> 基因沉默小鼠相继出现。另外, 一些 lncRNA 则被发现在发育过程中发挥重要调节作用, 如 *Fendrr*<sup>[36-37]</sup>、*Peril*<sup>[37]</sup>、*Mdgt*<sup>[37]</sup>、*Linc-Brn1b*<sup>[37]</sup>、*Hotdog (Hog)* 和 *Twin of Hotdog (Tog)*<sup>[37]</sup>。伴随着进化由简至繁, lncRNA 在基因组中所占比例显著增加。学者们因而推测 lncRNA 可能是与高等动物的某些复杂神经活动相关的重要调节因子。近期的 lncRNA 在体动物研究结果初步印证了这一推测。来自哈佛大学的 Rinn 研究组对 18 种小鼠 lncRNA 基因进行了敲除, 发现其中 *Spasm*<sup>-/-</sup> 和 *linc-p21*<sup>-/-</sup> 两种小鼠发生行为学改变, 该研究组正在对此进行深入的分析<sup>[37]</sup>。此外, 来自中国清华-北大生命科学联合中心的沈晓骅研究组在胚胎干细胞中, 对 *HOX* 基因座上的 lncRNA——*Halr1/Haunt* 进行了深入细致而全面的研究<sup>[44]</sup>, 有许多值得借鉴之处, 因此也在本文中作详细讨论。

在此阶段, 通过进一步对亚功能区域进行选择定位敲除, 学者们对 *Xist*<sup>[16-17]</sup> 和 *Kcnqlot1*<sup>[20]</sup> 等经典印记 lncRNA 基因进行了更为细致的在体研究(表1)。此外, 还有一些新的基因组印记 lncRNA, 如 *Jpx*<sup>[25]</sup> 和 *XACT*<sup>[45]</sup> 被发现。最后, 值得一提的是, 利用近两年新兴的 CRISPR-Cas9 技术已成功地获得了首例 lncRNA 基因敲除小鼠 *Rian*<sup>-/-</sup>(表1)<sup>[26]</sup>。以上这些研究不仅为 lncRNA 在体动物研究领域注入了新的活力, 而且为优化基因敲除方案和目的基因选择策略提供了宝贵经验(详细讨论见下文)。

## 2 lncRNA在体动物研究经典案例讨论

### 2.1 *Malat1/Neat2*

*MALATI/NEAT2* (metastasis associated lung adenocarcinoma transcript 1/ noncoding nuclear-enriched abundant transcript 2) 最初是在人非小细胞性肺癌患者组织中发现的, 在发生转移的患者中表达量显著上调<sup>[41]</sup>。Chess 研究组在一筛选细胞核中富集表达的 lncRNA 的研究中发现了 3 种 lncRNA, 其一为 *NEAT2*, 即 *MALATI*<sup>[46]</sup>。另外两个 lncRNA 为 *XIST* 和 *NEAT1*, 其中 *NEAT1* 是 *MALATI/NEAT2* 的邻近基因。研究发现, *MALATI/NEAT2* 在细胞核内富集于一种特殊的亚细胞结构核斑 (nuclear speckle) 中。该结构是许多 mRNA 前体剪切因子富集的区域。与这种特殊定位相符合, *MALATI/NEAT2* 能够与多种 mRNA 剪切因子结合并影响这些因子

的核斑定位。此外,来自数个独立研究组的工作发现, *MALAT1/NEAT2* 在细胞周期调节、细胞凋亡、RNA 代谢以及肿瘤转移等多种生理和病理过程中发挥着重要作用<sup>[3]</sup>。这些结果共同提示 *MALAT1/NEAT2* 是一个具有生物学功能的基因。基于这些发现,研究人员推测 *Malat1/Neat2* 基因敲除小鼠即便是能够生存的,也必定有显著发育异常。然而,令所有学者意外的是,在 2012 年,由 3 个独立小组分别通过不同的方法获得的 3 种 *Malat1/Neat2*<sup>-/-</sup> 小鼠和细胞均表现正常;进一步的研究发现,它们无论是在细胞生存、增殖、发育,还是在核斑的形成上均无显著异常<sup>[30-32]</sup>。在这 3 项研究工作中,仅发现 *Malat1/Neat2* 对极少数邻近基因(如 *Neat1*)的转录发挥调节作用。由于之前有关 *MALAT1/NEAT2* 重要生物学作用的研究大多数是在人源细胞中完成的,这些基于啮齿类动物的在体研究结果可能是由于人和鼠 *MALAT1/NEAT2* 间功能的差异而造成的。然而, Eissmann 等<sup>[32]</sup> 在进行在体动物研究的同时,利用锌指核酸酶技术(zinc finger nucleases)在人肝癌 HLE 和肺癌 A549 细胞上完全敲除 *MALAT1/NEAT2*,未观察到其对细胞周期和细胞生存等的影响。Nakagawa 等<sup>[30]</sup> 通过提前终止转录制备了 *Malat1/Neat2*<sup>-/-</sup> 小鼠,发现仅在一些特殊细胞,如小肠细胞中 *Neat1* 表达量下调。这些结果提示,在人与小鼠物种间功能差异之上, *MALAT1/NEAT2* 的作用可能还具有较高的细胞特异性,这些都可能是导致差异结果的原因。

然而,对 *MALAT1/NEAT2* 的研究并没有就此止步。近期, Liu 等<sup>[47]</sup> 研究发现, DNA 损伤刺激等信号使 *MALAT1/NEAT2* 表达水平发生了显著变化。Chang 研究组还发现通过结合在体肺癌模型, *Malat1/Neat2* 基因敲减小鼠的肿瘤迁移率显著降低<sup>[3]</sup>。最近来自冷泉港实验室的 Arun 等<sup>[48]</sup> 利用 *Malat1/Neat2* 纯和敲除雌鼠 (*Malat1*<sup>-/-</sup>)<sup>[49]</sup> 和 *MMTV-PyMT* 雄鼠 (C57Bl/6)(一种 luminal B 型乳腺癌小鼠肿瘤模型<sup>[50-52]</sup>) 杂交获得了 *MMTV-PyMT;Malat1*<sup>-/-</sup> 小鼠,与对照组相比, *MMTV-PyMT;Malat1*<sup>-/-</sup> 小鼠膀胱发生了充满囊性液体的肿瘤。与 Chang 研究组的研究结果一致, *MMTV-PyMT;Malat1*<sup>-/-</sup> 小鼠肿瘤向肺部迁移的速率显著降低。这些结果共同显示了 *MALAT1/NEAT2* 在肿瘤发生和迁移过程中扮演着重要的角色。

## 2.2 HOTAIR

*HOTAIR* (HOX transcript antisense RNA) 是

2007 年由斯坦福大学的 Chang 研究组发现的首例通过反式调控方式发挥作用的 lncRNA<sup>[53]</sup>。该研究发现 *HOTAIR* 位于 *HOXC* 基因座 *HOXC11* 与 *C12* 交界区域,通过促进组蛋白 H3K27 甲基化对 40 kb 之外的 *HOXD* 基因发挥反向调控作用。*HOTAIR* 的 5' 区域和 3' 区域可以分别与 PRC2 和 LSD1 结合,起到脚手架作用,对基因转录发挥抑制作用<sup>[2]</sup>。

为探索 *HOTAIR* 在发育过程中的作用,来自瑞士的 Duboule 研究组首先从已知的基因组信息中发掘出小鼠 *Hotair*<sup>[33]</sup>。人与小鼠 *Hotair* 的碱基序列保守性很低,但两者的空间位置分布高度保守。Duboule 研究组进而通过敲除 *HoxC* 基因座 (*HoxC*<sup>A</sup>) 的方法对 *Hotair* 的功能进行研究,其结果表明 *HoxC*<sup>A</sup> 小鼠生长发育正常,在 *HoxC*<sup>A</sup> 小鼠上不仅没有观察到基因组转录水平的改变,也没有检测到 *HoxD* 基因 H3K27 甲基化的影响。然而,在上述实验中, *HoxC*<sup>A</sup> 小鼠同时敲除了 8 个 *HoxC* 基因、2 个 microRNA 基因和一系列 lncRNA 基因。为排除来自其他基因的干扰,Chang 研究组采用仅敲除 *Hotair* 基因的方法制备了 *Hotair* 敲除 (*Hotair* KO) 小鼠<sup>[34]</sup>,发现与野生型 C57/BL6 小鼠相比, *Hotair* KO 小鼠发生了腰椎 L6 到 S1 的转化以及第 4 尾骨渗透性转化,并且超过一半的 *Hotair* KO 鼠发生了掌骨和腕骨的异常。此外, *Hotair* KO 鼠体内 H3K27 甲基化水平明显改变,基因组转录发生显著的去抑制 (de-repression),基本上重现人 *HOTAIR* 的功能。

## 2.3 Fendrr

在空间上, lncRNA 基因对应的 DNA 区域与蛋白质编码序列有着错综复杂的联系,有的与蛋白质编码基因重叠,有的具有增强子类似的功能,有的则与蛋白质编码基因共享一些重要的基因组元件。在对 lncRNA 基因进行编辑时,常有可能同时删除了一些重要的转录因子和染色体调节蛋白的 DNA 结合序列,或者改变了邻近基因的表达水平。这为正确解释 lncRNA 基因敲除动物的表型带来了诸多困难,需仔细甄别所观察到的表型确实是由 lncRNA 本身,而非其他间接因素引起的非特异性表型。

鼠 *Fendrr* (FOXF1 adjacent non-coding developmental regulatory RNA) 是一个位于第 8 号染色体上的基因间 lncRNA (long intergenic noncoding RNA, lincRNA),长度约为 2.4 kb,由 6~7 个外显子组成<sup>[36-37]</sup>。*Fendrr* 基因下游 1.3 kb 左右为蛋白质编码基因 *Foxfla*,两个基因共享一个双向的基因启动子。Herrmann 研究

组通过 RNA 原位杂交实验发现,小鼠 *Fendrr* 特异性表达于侧板中胚层背侧 (caudal lateral plate mesoderm, caudal LPM)<sup>[36]</sup>。该区域是心脏和体壁形成的关键部位。Herrmann 研究组进一步通过在 *Fendrr* 第 1 外显子处提前加入终止密码子的方法,获得了 *Fendrr* 基因沉默小鼠 *Fendrr*<sup>3xpA/3xpA</sup>, 这些小鼠由于心脏和体壁发育异常在胚胎期 E13.75 死亡。为排除 *Fendrr*<sup>3xpA/3xpA</sup> 小鼠的异常是由于 *Foxfla* 受到影响而引起的, Herrmann 研究组将含有失活 *Foxfla* 基因的 *Fendrr* BAC 克隆导入 *Fendrr*<sup>3xpA/3xpA</sup> 小鼠 ES 细胞, 得到发育正常的小鼠, 从而证明 *Fendrr*<sup>3xpA/3xpA</sup> 小鼠的表型是由于 *Fendrr* lncRNA 缺陷引起的。

与 Herrmann 研究组的工作略有不同, Rinn 研究组通过大片段敲除并以 *lacZ*-neo 取代的方法对 *Fendrr* 进行敲除, 获得的 *Fendrr*<sup>-/-</sup> 小鼠在出生后 24 h 内死亡<sup>[37]</sup>。Rinn 研究组通过 RNA-Seq 实验发现, *Fendrr* 的表达并不局限于侧板中胚层背侧, 而是在许多脏器中高表达, 肺脏尤为显著。与这种表达分布情况相吻合, *Fendrr*<sup>-/-</sup> 小鼠肺血管发育异常, 没有自主呼吸能力, 这可能正是 *Fendrr*<sup>-/-</sup> 小鼠围产期死亡的原因。为排除 *Foxfla* 的影响, Rinn 研究组对 *Fendrr*<sup>-/-</sup> 小鼠组织中 *Foxfla* 和其他 *Fendrr* 邻近基因的表达情况进行检测, 未发现这些基因的表达受到影响。这些结果也同样提示 *Fendrr*<sup>-/-</sup> 小鼠的表型是由于 *Fendrr* lncRNA 导致的。

#### 2.4 *Rian* (RNA imprinted and accumulated in nucleus)

通过同源重组配对获得基因敲除的传统技术效率较低。当需要大片段敲除时, 传统的基因敲除技术不仅耗时而且材料成本大大增加。近期发展起来的 CRISPR-Cas9 系统是一种能够靶向基因组中特异性 DNA 序列的基因编辑工具<sup>[39-40]</sup>。由于 CRISPR-Cas9 无需胚胎干细胞, 直接注射受精卵即可实现基因精确定点插入, 时间约为传统方法的十分之一, 材料成本也大大下降<sup>[39]</sup>。目前 CRISPR-Cas9 技术已成功地应用到多种在体动物研究中<sup>[40]</sup>。利用 CRISPR-Cas9 技术, 来自中国和英国的学者们通过合作成功地在小鼠体内对基因组印记 lncRNA 基因 *Rian* 进行了敲除<sup>[26]</sup>。*Rian* 是一个来自父本的印记基因, 位于鼠第 12 号染色体末端, 与另一个印记基因 *Meg3* (maternally expressed 3) 紧密相连, 其转录本特异性的定位在核内, 由 Hatada 等<sup>[54]</sup>首次发现。随后的敲除实验发现与 *Rian*<sup>+/- (pa)</sup> 和正常小鼠相比, *Rian* 的 6 个邻近基因在 *Rian*<sup>+/- (ma)</sup> 小鼠组织中的表达发生错调。

#### 2.5 *Halr1*

*Halr1* (Hoxa adjacent long noncoding RNA 1) 又名 *Haunt*, 位于 *HOXA* 上游 40 kb, 最初是由 Guttman 等<sup>[55]</sup>发现的, 在胚胎干细胞维持中发挥功能, 但作用机制不详。2013 年, Rinn 研究组曾报道 *Haunt* 能发挥顺式作用, 抑制 *HOXA* 的表达<sup>[56]</sup>。最近, 清华大学沈晓骅研究组利用 CRISPER-Cas9、基因干扰、3C 和 ChIRP 等技术, 在胚胎干细胞上对 *Haunt* 基因座 (DNA) 及其转录产生的 *Haunt* lncRNA 对 *HOXA* 基因调控的分子机制进行了深入细致的解析<sup>[44]</sup>。在该工作中, 作者发现 *Haunt* DNA 和 *Haunt* lncRNA 共同对 *HOXA* 基因的表达进行调控。*Haunt* 基因座上存在调控 *HOXA* 基因的增强子序列, 通过与 *HOXA* 基因启动子序列的结合促进 *HOXA* 基因的转录激活。而 *Haunt* lncRNA 则通过阻止 *Haunt* DNA 上增强子序列与 *HOXA* 基因的结合, 抑制 H3K27me3 的去甲基化, 抑制 *HOXA* 基因的表达, 起着制动器 (Brake) 的作用。为分别验证 *Haunt* DNA 和 *Haunt* lncRNA 的作用, 这项工作进行了巧妙的设计, 思路严谨, 工作细致而全面, 详细内容请参阅原文。

### 3 目的基因选择和敲除策略

如前所述, 许多 lncRNA 与蛋白质编码基因共享序列, 有些 lncRNA 还是 microRNA 等其他类型非编码 RNA 的前体。此外, 在基因组中还存在着一些有待发掘和确认的、具有蛋白质编码或合成短小肽链功能的 DNA 序列。如果在 lncRNA 基因编辑同时改变这些序列, 不仅有可能导致非特异性表型出现, 而且给实验结果的分析带来诸多困难。尽管还有许多内容需要进一步的完善, 目前解决这一难题可以从合理选择目的基因和结合每个 lncRNA 的个体情况制定相应的敲除策略两方面入手。在合理选择目的基因方面, Rinn 研究组在前人的工作基础上总结了一套行之有效的选择优化方案<sup>[37]</sup>。首先, 由于 lincRNA 位于基因间, 从空间位置上来看, 具有与其他基因交叉较少的优势。因此, 他们将目标锁定在 lincRNA 这一特殊类型的 lncRNA。然后, 结合多种计算分析和前期实验数据, 排除那些在序列上含有已知的或能够预测的其他任何基因的 lincRNA, 详细方法参阅原文。在敲除方法上, lncRNA 的研究基本上沿袭了编码基因常用的传统技术, 包括敲除基因启动子 (promoter deletion)、提前插入终止密码子 (premature transcriptional termination)、

反转 (inversion)、整个基因 / 基因座敲除 (locus/cluster deletion) 等。新兴的 CRISPR-Cas9 技术由于其高效低耗、时间短、用途广泛等优势有望成为加快推动 lncRNA 在体动物研究的强有力工具。根据目的基因的产生位点、功能方式选择合适的基因编辑方法是 lncRNA 在体研究的关键步骤, 详细讨论见近期综述<sup>[4]</sup>。简而言之, 如果通过 Rinn 研究组的方法在筛选目的基因时已经尽可能地排除了对其他基因序列的影响, 建议采用整个基因敲除或结合使用 lacZ 等标签基因取代的策略。然而, 绝大多数 lncRNA 不在 Rinn 研究组目的基因选择策略范畴之内, 在这种情况下, 则应尽可能减少敲除片段的长度, 以最大程度保留其他有功能的 DNA 元件。

#### 4 展望

综上所述, lncRNA 在体动物研究为从发育角度深入了解 lncRNA 的生理学功能提供了非常重要的信息。结合 lncRNA 本身的特性, 我们推测, 尽管有一些 lncRNA 是在生命体正常发育过程中发挥关键作用的基因, 而对于大多数 lncRNA 来说, 它们可能更多是在病理条件下发挥作用, 抑或是参与某些高级活动和与环境、饮食、情绪等相关的复杂性状的重要调节分子。现阶段, 如何合理选择目的基因和设计研究策略仍然是 lncRNA 在体动物研究的难点。由于 lncRNA 功能的复杂性, 可能需要通过建立不同基因编辑策略的动物模型而获得对某一特定 lncRNA 功能的全面认识。为正确解读实验结果, RNA rescue 实验将成为 lncRNA 在体动物研究必不可少的步骤。新技术、新方法和新的动物模型的发展决定着 lncRNA 在体动物研究的发展速度、广度和深度。我们期待在今后几年内, 将有更多 lncRNA 在体动物研究成果出现, 使我们对 lncRNA 的生理学和病理学功能有一个较为清晰的认识。

#### [参 考 文 献]

- [1] Wang X, Song X, Glass CK, et al. The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. *Cold Spring Harbor Perspect Biol*, 2011, 3: a003756
- [2] Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol*, 2011, 21: 354-61
- [3] Li LJ, Chang HY. Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends Cell Biol*, 2014, 24: 594-602
- [4] Bassett AR, Akhtar A, Barlow DP, et al. Considerations when investigating lncRNA function *in vivo*. *Elife*, 2014, 3: e03058
- [5] Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature*, 2012, 489: 101-8
- [6] Zieve G, Penman S. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell*, 1976, 8: 19-31
- [7] Brannan CI, Dees EC, Ingram RS, et al. The product of the *H19* gene may function as an RNA. *Mol Cell Biol*, 1990, 10: 28-36
- [8] Brown CJ, Ballabio A, Rupert JL, et al. A gene from the region of the human X-inactivation center is expressed exclusively from the inactive X-chromosome. *Nature*, 1991, 349: 38-44
- [9] Lyle R, Watanabe D, Vruchte D, et al. The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nat Genet*, 2000, 25: 19-21
- [10] Lee MP, DeBaun MR, Mitsuya K, et al. Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc Natl Acad Sci USA*, 1999, 96: 5203-28
- [11] Leighton PA, Ingram RS, Eggenschwiler J, et al. Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature*, 1995, 375: 34-9
- [12] Leighton PA, Saam JR, Ingram RS. An enhancer deletion affects both H19. *Gene Dev*, 1995, 9: 2079-89
- [13] Ripoche MA, Kress C, Poirier F, et al. Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Gene Dev*, 1997, 11:1596-604
- [14] Marahrens Y, Panning B, Dausman J, et al. *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. *Gene Dev*, 1997, 11:156-66
- [15] Csankovszki G, Panning B, Bates B, et al. Conditional deletion of *Xist* disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet*, 1999, 22: 323-4
- [16] Senner CE, Nesterova TB, Norton S, et al. Disruption of a conserved region of *Xist* exon 1 impairs *Xist* RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation. *Development*, 2011, 138: 1541-50
- [17] Yildirim E, Kirby JE, Brown DE, et al. *Xist* RNA is a potent suppressor of hematologic cancer in mice. *Cell*, 2013, 152: 727-42
- [18] Meller VH, Rattner BP. The *roX* genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J*, 2002, 21: 1084-91
- [19] Fitzpatrick GV, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of *KvDMR1*. *Nat Genet*, 2002, 32: 426-31
- [20] Mancini-Dinardo D, Steele SJ, Levorse JM, et al. Elongation of the *Kcnqlot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev*, 2006, 20: 1268-82
- [21] Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, 2002, 415: 810-3
- [22] Lee JT. Disruption of imprinted X inactivation by parent-

- of-origin effects at Tsix. *Cell*, 2000, 103: 17-27
- [23] Sado T, Wang Z, Sasaki H. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development*, 2001, 128: 1275-86
- [24] Anguera MC, Ma WY, Clift D, et al. *Tsix* produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet*, 2011, 7: e1002248
- [25] Tian D, Sun S, Lee JT. The long noncoding RNA, *Jpx*, is a molecular switch for X chromosome inactivation. *Cell*, 2010, 143: 390-403
- [26] Han J, Zhang J, Chen L, et al. Efficient *in vivo* deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol*, 2014, 11: 829-35
- [27] Zhong J, Chuang SC, Bianchi R, et al. BC1 regulation of metabotropic glutamate receptor-mediated neuronal excitability. *J Neurosci*, 2009, 29: 9977-86
- [28] 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
- [29] Nakagawa S, Naganuma T, Shioi G. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J Cell Biol*, 2011, 193: 31-9
- [30] Nakagawa S, Ip JY, Shioi G, et al. *Malat1* is not an essential component of nuclear speckles in mice. *RNA*, 2012, 18: 1487-99
- [31] Zhang B, Arun G, Mao YS, et al. The lncRNA *malat1* is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*, 2012, 2: 111-23
- [32] Eissmann M, Gutschner T, Hammerle M, et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol*, 2012, 9: 1076-87
- [33] Bickmore WA, Schorderet P, Duboule D. Structural and functional differences in the long non-coding RNA *hotair* in mouse and human. *PLoS Genet*, 2011, 7: e1002071
- [34] Li LJ, Liu B, Wapinski OL, et al. Targeted disruption of *hotair* leads to homeotic transformation and gene derepression. *Cell Rep*, 2013, 5: 3-12
- [35] Delpretti S, Montavon T, Leleu M, et al. Multiple enhancers regulate *hoxd* genes and the *hotdog* lncRNA during cecum budding. *Cell Rep*, 2013, 5: 137-50
- [36] 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. *Dev Cell*, 2013, 24: 206-14
- [37] Sauvageau M, Goff LA, Lodato S, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*, 2013, 2: e01749
- [38] Gomez JA, Wapinski OL, Yang YW, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- $\gamma$  locus. *Cell*, 2013, 152: 743-54
- [39] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 2014, 157: 1262-78
- [40] Niu Y, Shen B, Cui Y, et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*, 2014, 156: 836-43
- [41] Ji P, Diederichs S, Wang W, et al. MALAT-1, a novel noncoding RNA, and thymosin  $\beta$ 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 2003, 22: 8031-41
- [42] Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 2009, 10: 155-9
- [43] Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*, 2009, 23: 1494-504
- [44] Yin Y, Yan P, Lu J, et al. Opposing roles for the lncRNA *haunt* and its genomic locus in regulating *HOXA* gene activation during embryonic stem cell differentiation. *Cell Stem Cell*, 2015, 16: 504-16
- [45] Vallot C, Huret C, Lesecque Y, et al. *XACT*, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat Genet*, 2013, 45: 239-41
- [46] Hutchinson JN, Ensminger AW, Clemson CM, et al. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, 2007, 8: 39
- [47] Liu Q, Sun S, Yu W, et al. Altered expression of long non-coding RNAs during genotoxic stress-induced cell death in human glioma cells. *J Neurooncol*, 2015, 122: 283-92
- [48] Arun G, Diermeier S, Akerman M, et al. Differentiation of mammary tumors and reduction in metastasis upon *Malat1* lncRNA loss. *Genes Dev*, 2016, 30: 34-51
- [49] Zhang B, Arun G, Mao Y S, et al. The lncRNA *Malat1* is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*, 2012, 2: 111-23
- [50] Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol*, 1992, 12: 954-61
- [51] Lin EY, Jones JG, Li P, et al. Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am J Pathol*, 2003, 163: 2113-26
- [52] Herschkowitz JI, Simin K, Weigman VJ, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, 2007, 8: R76
- [53] Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 2007, 129: 1311-23
- [54] Hatada I, Morita S, Obata Y, et al. Identification of a new imprinted gene, *Rian*, on mouse chromosome 12 by fluorescent differential display screening. *J Biochem*, 2001, 130: 187-90
- [55] Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 2011, 477: 295-300
- [56] Maamar H, Cabili MN, Rinn J, et al. *linc-HOXA1* is a noncoding RNA that represses *Hoxa1* transcription in cis. *Genes Dev*, 2013, 27: 1260-71