

DOI:10.16781/j.0258-879x.2017.10.1304

长链非编码 RNA 的保守性及其在非模式生物长链非编码 RNA 筛选中的应用

姜贵先¹, 罗溪², 张露露³, 刘青³, 肖良^{4*}

1. 第二军医大学学员旅学员 2 队, 上海 200433
2. 第二军医大学学员旅学员 8 队, 上海 200433
3. 山西农业大学动物科技学院, 晋中 030801
4. 第二军医大学海军医学系海洋生物技术教研室, 上海 200433

[摘要] 长链非编码 RNA(long noncoding RNA, lncRNA)的保守性表现在一级结构、空间结构、转录位置、剪接模式及组织分布等方面,是目前 lncRNA 研究的热点和难点。不断深入的 lncRNA 保守性研究可应用于参考基因组相对匮乏的非模式生物 lncRNA 的筛选过程,并且极大地提升了非模式生物 lncRNA 数据库建立的完整性和准确性。借助针对开放阅读框长度、密码子分布与出现频率、功能性结构域等保守信息开发而来的 lncRNA 筛选工具或流程如 CPC、PLAR(pipeline for lncRNA annotation from RNA-seq data)等,已成为目前非模式生物 lncRNA 的筛选及其参考数据库构建的新策略。本文就 lncRNA 的保守性及其在非模式生物 lncRNA 筛选中的应用作一综述,并简要介绍了一种运用其保守性的筛选方法——PLAR。

[关键词] 长链非编码 RNA;保守性;非模式生物;筛选

[中图分类号] R 342.2 **[文献标志码]** A **[文章编号]** 0258-879X(2017)10-1304-07

Conservation of long noncoding RNA and its application in screening long noncoding RNA in non-model organisms

JIANG Gui-xian¹, LUO Xi², ZHANG Lu-lu³, LIU Qing³, XIAO Liang^{4*}

1. The Second Student Team, Student Brigade, Second Military Medical University, Shanghai 200433, China
2. The Eighth Student Team, Student Brigade, Second Military Medical University, Shanghai 200433, China
3. College of Animal and Technology, Shanxi Agricultural University, Jinzhong 030801, Shanxi, China
4. Department of Marine Biotechnology, Faculty of Naval Medicine, Second Military Medical University, Shanghai 200433, China

[Abstract] The conservation of long noncoding RNA (lncRNA), including primary structure, spatial structure, transcriptional position, splicing mode and tissue distribution, is research focus of lncRNA research. In-depth studies of lncRNA conservation can be applied to the screening process of lncRNA in non-modal organisms, which lack adequate reference genome databases, greatly enhancing the integrity and accuracy of lncRNA databases of non-modal organisms. LncRNA screening tools or processes such as CPC and pipeline for lncRNA annotation from RNA-seq data (PLAR), which have been developed by length of open reading frame, distribution and frequency of codon, functional domain and other conservative information, are new strategies for the screening of lncRNA and construction of reference database of non-model organisms. In this review, we summarized the conservation of lncRNA and its application in screening lncRNA in non-model organisms, and briefly introduced PLAR, a new screening method using the conservation of lncRNA.

[Key words] long noncoding RNA; conservation; non-model organisms; screening

[Acad J Sec Mil Med Univ, 2017, 38(10): 1304-1310]

[收稿日期] 2017-05-30 **[接受日期]** 2017-09-11

[基金项目] 国家自然科学基金(81470518),上海市浦江人才项目(16PJ1411000). Supported by National Natural Science Foundation of China (81470518) and Shanghai Pujiang Talent Program (16PJ1411000).

[作者简介] 姜贵先,第二军医大学临床医学专业五年制 2015 级学员。E-mail: jiangdaxian@hotmail.com

* 通信作者 (Corresponding author). Tel: 021-81871129, E-mail: hormat830713@hotmail.com

长链非编码 RNA (long noncoding RNA, lncRNA) 是一类长度大于 200 nt 的非编码 RNA, 在个体生长发育、疾病发生及发展等生命过程中发挥重要作用^[1]。通过比较模式生物间、模式生物与非模式生物间甚至非模式生物间 lncRNA 的序列、转录位置、空间结构以及生物学功能等特征分析其保守性, 并应用于新 lncRNA 的筛选与鉴定是近年来 lncRNA 研究的新策略^[2]。

物种 lncRNA 参考数据库的构建是开展 lncRNA 相关研究的前提条件, 其主要步骤包括样品测序、序列组装和数据库比对注释等。模式生物由于具有更为完备的基因组、转录组以及 lncRNA 自身参考数据库, 新 lncRNA 筛选与鉴定研究的策略相对完善且固定^[3]。而非模式生物由于可供参考的数据库信息有限, 其 lncRNA 的筛选、鉴定及其所构建的参考数据库易出现注释不全面、不准确甚至错误等问题^[4]。根据 lncRNA 的保守性特征开发生物信息学分析工具, 通过多层次筛选可有效提高非模式生物 lncRNA 参考数据库构建的准确性, 更好地推动新 lncRNA 的鉴定与功能研究进程^[5]。本文就 lncRNA 保守性的研究现状及其如何运用于非模式生物 lncRNA 的筛选作一综述。

1 lncRNA 的保守性

lncRNA 的保守性一般表现在一级结构、转录位置、二级结构以及功能与组织特异性等。借鉴蛋白质^[6]和微 RNA (miRNA)^[7] 的保守性在序列辨认、靶点预测和生物学效应研究等方面的应用, 深入研究 lncRNA 的保守性有助于解决序列比对、分析和预测等实际问题。

1.1 序列保守性 与信使 RNA (mRNA) 类似, 新合成的 lncRNA 含有 5' 端 7-甲基鸟嘌呤帽子、3' 端 poly A 尾以及开放阅读框 (opening reading frame, ORF) 等基本结构, 需要经过剪接加工成为成熟的 lncRNA^[8]。通过物种间同源基因的比较分析发现 lncRNA 的平均长度大约为编码基因长度的 1/5 (1 000 nt 左右), 有很大概率含有多个 50~100 nt 的 ORFs, 如有研究发现脊椎动物基因间的 lncRNA 平均含有 2~3 个外显子^[5,9]。研究选取高表达、经过精确注释的 lncRNA 构建数据库, 分析发现 lncRNA 外显子的保守性低于 mRNA 外显子的保守性^[10-12], 但高于编码基因内含子、基因间序列

(intergenic regions)^[11,13-15] 以及 lncRNA 自身内含子^[13] 的保守性。就不同物种而言, lncRNA 的保守性非常低, 如研究发现人与小鼠间 lncRNA 的保守性仅 20%, 而人和鱼之间甚至降低至仅 5%^[5]。目前已发现的 lncRNA 基因座超过 58 000 个, 但实际鉴定出的具有生物学功能的 lncRNA 却十分有限, 其数量远不及蛋白质, 可能与 lncRNA 具有较低的序列保守性有关。

1.2 空间结构的保守性 lncRNA 的空间结构对其生物学功能具有重要作用。研究表明, 生物体内的 lncRNA 往往比 mRNA 具有更多的二级结构^[16-17], 已有实验证实部分 lncRNA 中的局部区域具有特定的二级结构, 如肺癌转移相关转录本 1 (metastasis-associated lung adenocarcinoma transcript 1, MALAT1) 3' 末端的三螺旋 (triple helix) 结构具有类似增强子的作用^[18], 心肌肥厚相关表观遗传调节子 (cardiac-hypertrophy-associated epigenetic regulator, CHAER) 在 5' 端的双四环 (bitetra-loop) 结构可能是与 *Zeste* 基因增强子同源物 2 (enhancer of zeste homolog 2, EZH2) 结合的必要条件^[19] 等。需要注意的是, 具有相同二级结构 lncRNA 的基因在物种间未必是同源的^[5]。目前尚无 lncRNA 三级、四级结构的研究报道。近期开发的 dChIRP (domain-specific ChIRP)^[20] 技术可以在天然环境下剖析 lncRNA 不同结构域的功能, 如研究揭示果蝇体内一种雄性果蝇剂量补偿效应所需的 lncRNA roX1 的三指手掌结构 (three-fingered hand) 可与染色质相互作用, 是未来探索 lncRNA 高级结构对其生物学功能影响的可行方法之一。

1.3 转录位置的保守性 lncRNA 转录位置的保守性是指在直系同源序列中 lncRNA 与编码基因或其他区域相对位置的固定^[21], 这提示 lncRNA 对编码基因存在如沉默作用等某种特定的调控。研究发现非蛋白质编码 RNA IGF2R 反义链 (antisense of IGF2R non-protein coding RNA, AIRN) 的基因序列和编码基因胰岛素样生长因子 2 型受体 (insulin-like growth factor 2 receptor, IGF2R) 基因的启动子之间存在部分重叠, 重叠部分转录出的 RNA 可沉默 IGF2R, 而剩余大小为 118 kb 的非重叠 AIRN 序列则与基因沉默无关^[22]。同样, 不同哺乳动物间的 lncRNA 浆细胞瘤转化迁移基因 1 (plasmacytoma variant translocation 1, PVT1) 仅在

第1个及中间2个外显子上存在保守性,但它们都位于 *Myc* 编码基因的下游^[5]。转录位置的保守性是极其容易忽略的,符合这一性质的 lncRNA 在序列上往往并不保守,且由于其具有功能的部分片段长度较短或其转录需要特定的条件,从而导致这类保守性难以被发现。

1.4 转录和剪接模式的保守性 比较 lncRNA 的外显子是在 RNA 层面研究物种 lncRNA 进化保守性的一种常规思路^[5,10,23]。但研究常发现在两亲缘关系很近的物种中存在 lncRNA 特异性表达于单一物种的现象,提示有必要进一步从 DNA 层面比较 lncRNA 的基因座,分析 lncRNA 转录和剪接模式的保守性。Hezroni 等^[5]通过对 17 种脊椎动物的 lncRNA 研究发现 >70% 的 lncRNA 基因座已存在于 5 000 万年前的生物体中。但与非灵长类动物相比,人类仅保留了大约 20% 的 lncRNA 剪接过程^[23],说明 lncRNA 具有非常高的进化效率。与其他依赖保守 DNA 或成熟 RNA 序列的蛋白质或 miRNA 相比,很多转录 lncRNA 的基因座会在进化中逐渐改变或丢失,而保留至今的 lncRNA 的内外显子的序列和结构往往较之前发生了较大的变化。这一方面可能是由于本身具有功能的 lncRNA 数量并不多,容易在进化过程中丢失^[4];另一方面则可能是由于 lncRNA 需要与其他元件结合才能发挥作用,如与 miRNA 或蛋白结合发挥内源性竞争作用,未必表现出明显的保守性^[24]。

1.5 组织和细胞特异性 lncRNA 在表观遗传调控、细胞周期调控和细胞分化调控等众多生命活动中发挥重要作用。高度保守的 lncRNA 往往具有组织特异性,常存在于脑、睾丸等体内重要组织器官中^[5,23,25],发挥调控胚胎发育、精子生成以及突触传递等重要的生理功能^[10]。Ulitsky 等^[26]在斑马鱼中发现了 29 条与哺乳动物同源的 lncRNA,其中 *cyrano* 和 *megamind* 主要分布于神经系统中,可能与人类胚胎期大脑发育有关。但总体而言,lncRNA 分布的组织特异性并不高,在不同生物体内的相同特定细胞中,同源 lncRNA 的数量非常有限。Kutter 等^[27]研究发现小鼠肝脏与大鼠肝脏有 60% 的 lncRNA 同源,而人肝脏中 lncRNA 的同源性降低至 27%。同样,在人和鼠眼(视锥、视杆和视网膜细胞)^[28]、多能干细胞^[29]以及胰岛朗格汉斯细胞^[30]

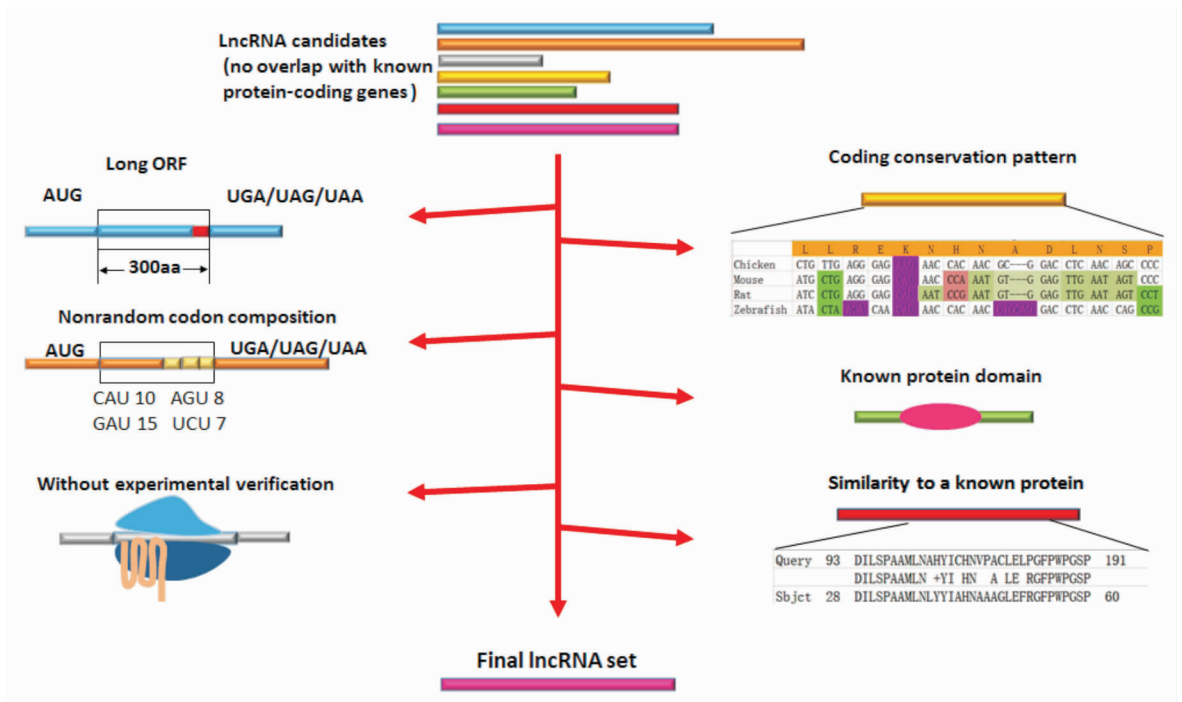
中 lncRNA 的同源性结果类似。

1.6 功能保守性 尽管 lncRNA 数量众多,但仅少数具有明确的生物学功能。从保守性入手对 lncRNA 的序列、结构、加工过程及空间分布等进行比较,最终目的在于揭示 lncRNA 可能存在的保守功能,如序列上十分保守的 Uc. 283 + A 能结合 miR-195 前体并下调 miR-195 的表达^[31]。lncRNA 的保守功能可以体现在多个方面,如具有相似的分子功能、与相同靶基因结合甚至代替其他生物中的同源序列并发挥相似作用,有研究人员发现其他生物中的 roX 可以在缺失 roX 的果蝇体内与其相应位点结合^[32]。由于目前发现的保守 lncRNA 数量较少,以及 lncRNA 活体实验技术的不成熟,发现具有保守功能的 lncRNA 并对其实验验证仍是深入研究 lncRNA 功能保守性的难点。

2 保守性在非模式生物 lncRNA 筛选中的应用

非模式生物往往缺少相对应的基因组参考数据库,直接将组装好的序列与模式生物基因组比对,进行 lncRNA 的筛选会导致比对不精确或比对不全,有时甚至会注释出错误信息。现行的主要处理办法一是选择亲缘较近的模式生物,以求注释得到的信息更加精确,二是与数据库中所有物种进行匹配,以获得尽可能多的注释。借助不断涌现的 lncRNA 保守性研究成果与 mRNA 的相互比较,采用正向选择(针对 lncRNA)和反向排除(针对 mRNA)的方法,不断开发新的软件工具并进行优化和完善^[2]成为目前各实验室开展非模式生物 lncRNA 筛选的重要方式^[5,10,23,33]。图 1 显示了 lncRNA 保守性在非模式生物 lncRNA 筛选中的应用^[21]。

2.1 ORF 长度 ORF 或外显子的长度大于 300 nt 是大多数 mRNA 的特征^[9]。采用 CPAT^[34] 和 iSeeRNA^[35] 等工具检测 ORF 的数量和长度并反向排除含有大于 300 nt ORF 的序列是 lncRNA 筛选过程中的简单步骤。显然,仅根据 ORF 长度进行 lncRNA 的筛选是十分有限的^[9],如目前研究较深入的 *Xist*、*Hotair*、*Meg3* 等 lncRNA 的部分外显子长度均大于 300 nt。一般 ORF 长度在 lncRNA 筛选中并不会单独应用,但由于其操作简便,已成为 lncRNA 筛选中普遍使用的方法。

图 1 lncRNA 保守性在非模式生物 lncRNA 筛选中的应用示意图^[21]Fig 1 Diagram of application of lncRNA conservation in screening lncRNA in non-model organisms^[21]

lncRNA: Long noncoding RNA; ORF: Opening reading frame

2.2 核苷酸或密码子的出现频率与分布 lncRNA ORF 的核苷酸突变频率高,位置往往呈随机分布,而编码 mRNA 的 ORF 这一特征则正好相反,这与长期的生物进化过程需要维持编码蛋白的功能稳定性是一致的。这种由生物选择压力(selective pressure)导致的 lncRNA 和 mRNA ORF 之间的核苷酸突变频率差异表明,可以通过比较 ORF 内不同位置上的碱基构成或测试 ORF 内碱基突变对其生物学功能的影响来区分 lncRNA 与 mRNA。由于突变频率分析简单、计算快捷,其也已成为区分筛选 lncRNA 的常用特征参数之一,常见工具有 CPC^[36]、CONC^[37] 和 CPAT^[34]。其中 CPC 和 CONC 对比分析的是单个密码子及其表达的氨基酸的突变频率,而 CPAT 则是观察每个碱基在 ORF 中以及每 2 个密码子在整条序列中的出现频率。

2.3 功能性结构域 以 lncRNA ORF 推测翻译的蛋白质虽然也存在小范围、不常见的保守结构域,但其数量和结构种类远少于 mRNA 翻译的蛋白质所含有的结构域。HMMER (<http://www.hmm.org/>) 可用于检验转录产物、分析其编码的蛋白质序列并计算其潜在的结构域。可参考的数据库如 Pfam^[38] (<http://pfam.xfam.org/>) 运用多序列对比和隐马尔可夫模型(hidden Markov model, HMM)收录了大量蛋白家族,其条目已超过 16 000。GenBank 和 Ensembl 收录的蛋白质更多,但与

Pfam 收录蛋白质必须经实验验证相比,其存在大量仅由计算机推测获取的蛋白质序列,导致 lncRNA 假阴性的概率增加。

2.4 实验检验 在完成生物信息学筛选后,针对感兴趣 lncRNA 需进一步实验验证,其验证核心在于是否会翻译出蛋白质。由于多数转录本仅含有个别长的保守 ORF,因此可以采用移码突变(frameshift mutations)的方式来干扰其翻译成多肽并探讨其对 RNA 功能的影响,从而判断其是否为 lncRNA^[26,39]。

ORF 的长度、密码子的组成与分布、序列所翻译蛋白质的结构域与相似度等特征都是 lncRNA 保守性运用于非模式生物筛选中的重要参数指标。非模式生物缺少完整、准确的基因组信息,而保守程度的算法也不局限于序列的保守性,还包括空间结构、转录位置以及剪接方式等,导致目前已建立的 lncRNA 保守性算法难以完全涵盖,实现全面、准确和高效的非模式生物 lncRNA 的筛选仍有待进一步优化和完善。

3 PLAR (pipeline for lncRNA annotation from RNA-seq data) 流程在 lncRNA 筛选中的应用

以色列 Weizmann 研究所的 Ulitsky 教授及其实验室是 lncRNA 保守性研究^[4-5,21,26,40]的顶尖团队,他们构建了一套称为 PLAR^[5]的包括非模式生物在内的完整的 lncRNA 筛选方法(图 2)。PLAR

的简要过程如下:(1)转录组信息搜集,即通过测序或借助公共数据库获取研究对象完整的转录组信息;(2)转录本的组装,即使用 TopHat2 将测序所获得的 RNA 序列借助 Ensembl 和 RefSeq 数据库配对到基因组中,配对结果用 Cufflinks 进一步组装成转录模型,对于碎片化的短序列则使用 Trinity 进行从头 (*De novo*) 组装;(3)信息注释,即采用 CuffMerge 软件,以 Ensembl 为参考数据库对所建立的转录模型进行注释;(4)lncRNA 的筛选,即利用 CPC^[36]、HMMER^[41] 和 RNAcode^[42] 等方法或数据库以及 pseudogene(<http://pseudogene.org>) 排除与假基因重叠的转录本,并去除低表达的 RNA 序列;(5)排除有编码潜力的转录本,即设置一定的参数排除与蛋白质编码基因重叠或相近的序列,如将与编码基因外显子重叠超过 50% 的单外显子转录本排除;(6)将剩余 lncRNA 分类,包括基因区间的长链非编码 RNA (long intergenic noncoding RNA, lincRNA)、antisense lncRNA 以及 small RNA hosts 等。PLAR 的详细过程、使用软件及参数、筛选数据库等信息可参见网站 <http://webhome.weizmann.ac.il/home/igoru/PLAR>。

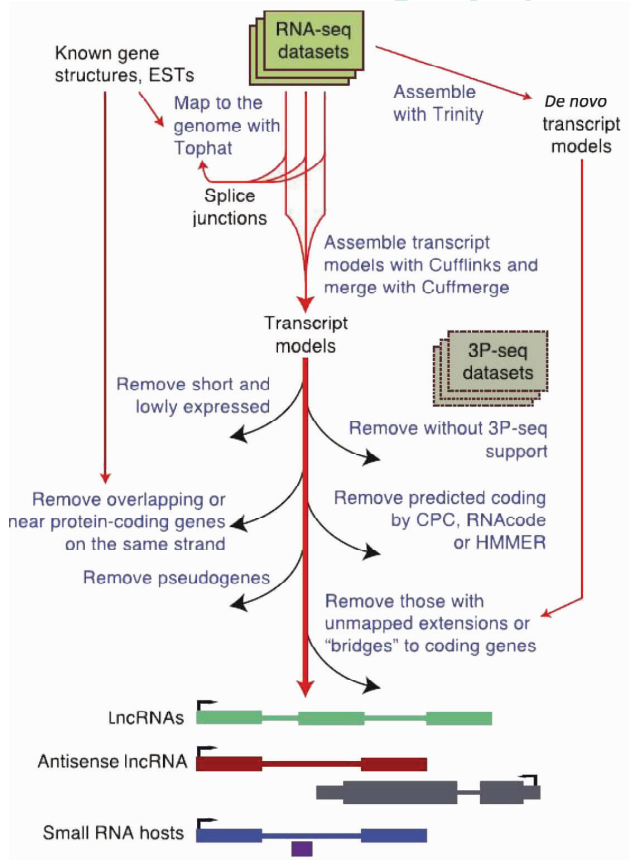


图 2 PLAR 流程示意图^[5]

Fig 2 Diagram of PLAR flow^[5]

PLAR: Pipeline for lncRNA annotation from RNA-seq data;
lncRNA: Long non-coding RNA; ESTs: Expressed sequence tags

目前已建立的 PLAR 数据库包括人、恒河猴、斑马鱼、海胆、姥鲨等 17 种非模式生物及常规模式生物,并发现这些生物中的同源 lincRNA 在 5' 端外显子中具有小段保守序列,且部分与转录因子重叠;在哺乳动物中预测超过 1 000 条具有保守功能的人 lincRNA,在非哺乳动物中也具有数百条,这些序列能在基因结构大幅度改变的情况下仅需特定片段中的一小段即可行使其功能,这大大扩展了 lncRNA 保守性的相关研究。PLAR 在常规 lncRNA 建库策略的基础上加以改进,借助公认的数据库获取最全面的转录组信息(正向选择),通过保守性多层次筛选编码 RNA(反向排除),在现有技术条件下可以获得较为准确、完整、高质量的 lncRNA 数据库,具有较高的推广价值,是目前运用保守性筛选 lncRNA 的代表性方法之一,可将其应用于几乎所有物种。

4 小结

保守性是目前 lncRNA 研究的热点和难点。尽管研究发现 lncRNA 在一级结构、空间结构、转录与剪接模式以及分布和功能等方面都存在一定的保守性,但其保守性明显低于 mRNA,且需更多的实验数据支撑。在缺少对应基因组数据库的情况下,根据 lncRNA 与 mRNA 的保守性同时采用正向选择和反向排除的方法,可以筛选出非模式生物的 lncRNA 并且构建相应的参考数据库。但由于该过程受到 lncRNA 和编码基因重叠程度、分析软件、参考数据库以及算法等多种因素的影响,其完整性和准确性在不同实验室间存在着明显差异。随着 lncRNA 保守性研究的不断深入和生物信息学筛选工具的不断完善,所构建的非模式生物 lncRNA 参考数据库将更全面和准确,也将更具参考价值。

[参考文献]

[1] CHEN L L. Linking long noncoding RNA localization and function[J]. Trends Biochem Sci, 2016, 41: 761-772.
[2] HOUSMAN G, ULITSKY I. Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs [J]. Biochim Biophys Acta, 2016, 1859: 31-40.
[3] IYER M K, NIKNAFS Y S, MALIK R, SINGHAL

- U, SAHU A, HOSONO Y, et al. The landscape of long noncoding RNAs in the human transcriptome[J]. *Nat Genet*, 2015, 47: 199-208.
- [4] ULITSKY I. Evolution to the rescue; using comparative genomics to understand long non-coding RNAs[J]. *Nat Rev Genet*, 2016, 17: 601-614.
- [5] HEZRONI H, KOPPSTEIN D, SCHWARTZ M G, AVRUTIN A, BARTEL D P, ULITSKY I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species[J]. *Cell Rep*, 2015, 11: 1110-1122.
- [6] BRAWAND D, SOUMILLON M, NECSULEA A, JULIEN P, CSÁRDI G, HARRIGAN P, et al. The evolution of gene expression levels in mammalian organs[J]. *Nature*, 2011, 478: 343-348.
- [7] BEREZIKOV E. Evolution of microRNA diversity and regulation in animals[J]. *Nat Rev Genet*, 2011, 12: 846-860.
- [8] QUINN J J, CHANG H Y. Unique features of long non-coding RNA biogenesis and function[J]. *Nat Rev Genet*, 2016, 17: 47-62.
- [9] DINGER M E, PANG K C, MERCER T R, MATTICK J S. Differentiating protein-coding and noncoding RNA; challenges and ambiguities[J/OL]. *PLoS Comput Biol*, 2008, 4: e1000176. doi: 10.1371/journal.pcbi.1000176.
- [10] NECSULEA A, SOUMILLON M, WARNEFORS M, LIECHTI A, DAISH T, ZELLER U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods[J]. *Nature*, 2014, 505: 635-640.
- [11] GUTTMAN M, AMIT I, GARBER M, FRENCH C, LIN M F, FELDSER D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals[J]. *Nature*, 2009, 458: 223-227.
- [12] ZHANG Y C, LIAO J Y, LI Z Y, YU Y, ZHANG J P, LI Q F, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice[J/OL]. *Genome Biol*, 2014, 15: 512. doi:10.1186/513059-014-0512-1.
- [13] CABILI M N, TRAPNELL C, GOFF L, KOZIOL M, TAZON-VEGA B, REGEV A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses[J]. *Genes Dev*, 2011, 25: 1915-1927.
- [14] MARQUES A C, PONTING C P. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness [J]. *Genome Biol*, 2009, 10: R124.
- [15] PONJAVIC J, PONTING C P, LUNTER G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs [J]. *Genome Res*, 2007, 17: 556-565.
- [16] YANG J R, ZHANG J. Human long noncoding RNAs are substantially less folded than messenger RNAs[J]. *Mol Biol Evol*, 2015, 32: 970-977.
- [17] SPITALE R C, FLYNN R A, ZHANG Q C, CRISALLI P, LEE B, JUNG J W, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms [J]. *Nature*, 2015, 519: 486-490.
- [18] WILUSZ J E, BAPTISTE C K, LU L Y, KUHN C D, JOSHUA-TOR L, SHARP P A. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails[J]. *Genes Dev*, 2012, 26: 2392-2407.
- [19] WANG Z, ZHANG X J, JI Y X, ZHANG P, DENG K Q, GONG J, et al. The long noncoding RNA *Chaer* defines an epigenetic checkpoint in cardiac hypertrophy [J]. *Nat Med*, 2016, 22: 1131-1139.
- [20] QUINN J J, ILIK I A, QU K, GEORGIEV P, CHU C, AKHTAR A, et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification [J]. *Nat Biotechnol*, 2014, 32: 933-940.
- [21] ULITSKY I, BARTEL D P. lincRNAs: genomics, evolution, and mechanisms[J]. *Cell*, 2013, 154: 26-46.
- [22] LATOS P A, PAULER F M, KOERNER M V, ŞENERGIN H B, HUDSON Q J, STOCSTITS R R, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing [J]. *Science*, 2012, 338: 1469-1472.
- [23] WASHIETL S, KELLIS M, GARBER M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals[J]. *Genome Res*, 2014, 24: 616-628.
- [24] TAN J Y, SIREY T, HONTI F, GRAHAM B, PIOVESAN A, MERKENSCHLAGER M, et al. Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells [J]. *Genome Res*, 2015, 25: 655-666.

- [25] CHODROFF R A, GOODSTADT L, SIREY T M, OLIVER P L, DAVIES K E, GREEN E D, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes[J/OL]. *Genome Biol*, 2010, 11: R72. doi:10.1186/gb-2010-11-7-r72.
- [26] ULITSKY I, SHKUMATAVA A, JAN C H, SIVE H, BARTEL D P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution[J]. *Cell*, 2011, 147: 1537-1550.
- [27] KUTTER C, WATT S, STEFFLOVA K, WILSON M D, GONCALVES A, PONTING C P, et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression [J/OL]. *PLoS Genet*, 2012, 8: e1002841. doi: 10.1371/journal.pgen.1002841.
- [28] MUSTAFI D, KEVANY B M, BAI X, MAEDA T, SEARS J E, KHALIL A M, et al. Evolutionarily conserved long intergenic non-coding RNAs in the eye [J]. *Hum Mol Genet*, 2013, 22: 2992-3002.
- [29] CHEN J, SHISHKIN A A, ZHU X, KADRI S, MAZA I, GUTTMAN M, et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs[J]. *Genome Biol*, 2016, 17: 19.
- [30] MORÁN I, AKERMAN I, VAN DE BUNT M, XIE R, BENAZRA M, NAMMO T, et al. Human β cell transcriptome analysis uncovers lincRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes[J]. *Cell Metab*, 2012, 16: 435-448.
- [31] LIZ J, PORTELA A, SOLER M, GÓMEZ A, LING H, MICHLEWSKI G, et al. Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region[J]. *Mol Cell*, 2014, 55: 138-147.
- [32] QUINN J J, ZHANG Q C, GEORGIEV P, ILIK I A, AKHTAR A, CHANG H Y, et al. Rapid evolutionary turnover underlies conserved lincRNA-genome interactions [J]. *Genes Dev*, 2016, 30: 191-207.
- [33] BU D, LUO H, JIAO F, FANG S, TAN C, LIU Z, et al. Evolutionary annotation of conserved long non-coding RNAs in major mammalian species [J]. *Sci China Life Sci*, 2015, 58: 787-798.
- [34] WANG L, PARK H J, DASARI S, WANG S, KOCHER J P, LI W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model[J/OL]. *Nucleic Acids Res*, 2013, 41: e74. doi: 10.1093/nar/gkt006.
- [35] SUN K, CHEN X, JIANG P, SONG X, WANG H, SUN H. iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data[J]. *BMC Genomics*, 2013, 14(Suppl 2): S7.
- [36] KONG L, ZHANG Y, YE Z Q, LIU X Q, ZHAO S Q, WEI L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine[J/OL]. *Nucleic Acids Res*, 2007, 35(Web Server issue): W345-W349. doi: 10.1093/nar/gkm391.
- [37] LIU J, GOUGH J, ROST B. Distinguishing protein-coding from non-coding RNAs through support vector machines[J/OL]. *PLoS Genet*, 2006, 2: e29. doi: 10.1371/journal.pgen.0020029.
- [38] FINN R D, BATEMAN A, CLEMENTS J, COGGILL P, EBERHARDT R Y, EDDY S R, et al. Pfam: the protein families database[J]. *Nucleic Acids Res*, 2014, 42: D222-D230.
- [39] HU W, YUAN B, FLYGARE J, LODISH H F. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation [J]. *Genes Dev*, 2011, 25: 2573-2578.
- [40] AUYEUNG V C, ULITSKY I, McGEARY S E, BARTEL D P. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing[J]. *Cell*, 2013, 152: 844-858.
- [41] FINN R D, CLEMENTS J, ARNDT W, MILLER B L, WHEELER T J, SCHREIBER F, et al. HMMER web server: 2015 update[J/OL]. *Nucleic Acids Res*, 2015, 43(W1): W30-W38. doi: 10.1093/nar/gkv397.
- [42] WASHIETL S, FINDEISS S, MÜLLER S A, KALKHOF S, VON BERGEN M, HOFACKER I L, et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data [J]. *RNA*, 2011, 17: 578-594.