长链非编码RNA的常用研究技术和方法

王 卉 张玲玲 王冰蕊 任思蕊 高 洁 佟静媛 刘金花* 石莉红*

(中国医学科学院、北京协和医学院血液学研究所、血液病医院, 实验血液学国家重点实验室, 天津 300020)

摘要 长链非编码RNA(long noncoding RNA, lncRNA)在胚胎发育、谱系分化、基因印迹、疾病发生等过程中都发挥着非常重要的调控作用, 是转录调控的重要分支及热点。随着转录组学高通量测序技术的日臻完善, 越来越多的lncRNA被发现。为了全面解析lncRNA的作用及调控机制, 该综述从lncRNA的发现、表达、细胞定位、结构、功能、机制研究等不同方面总结了当前常用的研究技术、方法及它们的最新进展, 为其深入研究提供了基础和线索。

关键词 lncRNA; 高通量转录组测序技术; CRISPR/Cas9技术; ChIRP; 染色体构象捕获技术; lncRNA数据库

The Mini-review of Techniques and Methods Used for Long Noncoding RNA Studies

Wang Hui, Zhang Lingling, Wang Bingrui, Ren Sirui, Gao Jie, Tong Jingyuan, Liu Jinhua*, Shi Lihong*

(*State Key Laboratory of Experimental Hematology, Institute of Hematology, Blood Disease Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China*)

Abstract Long noncoding RNA (lncRNA), a novel class of noncoding RNA molecules longer than 200 nucleotides, are identified largely due to the rapid development of the next generation RNA sequencing technology. Recent studies reveal that lncRNAs play essential roles in a broad range of biological processes, including the embryonic development, lineage differentiation, genomic imprinting and disease pathology, etc. Given the necessary functions of lncRNAs, it is important to summary the techniques and methods currently used for lncRNA studies, which will shed more light on future lncRNA studies. Therefore, in this study we review the frequently-used mothods for the identification, expression, function, as well as the regulatory mechanism of lncRNAs, and update their most recent development.

Keywords lncRNA; RNA-Seq; CRISPR/Cas9; ChIRP; chromosome conformation capture; lncRNA databases

人类基因组中仅2%的转录子可以编码蛋白质, 但至少80%的基因组可以被转录[1]。在这些非编码蛋白质的转录子中, 绝大部分属于长链非编码RNA (long noncoding RNA, lncRNA), 即长度超过200 bp, 开放阅读框(open reading frame, ORF)小于300 bp的转录子。研究表明, lncRNA在胚胎发育、谱系分化、基因印迹、疾病发生等过程中都发挥着非常重要的调控作用, 是转录调控网络的关键组成部分[2-5]。随着对lncRNA生物功能认识的不断深入, 它的研究也逐渐成为生命科学领域转录调控的重要分支和热点。为了更加系统地认识lncRNA在正常组织发育分化以及异常病变中的作用, 我们总结了当前研究lncRNA的一些常用方法, 为lncRNA的深入研究提供参考。

收稿日期: 2016-01-13 接受日期: 2016-05-06
国家自然科学基金(批准号: 31471291)资助的课题
*通讯作者。Tel: 022-23909446, E-mail: liufeng198308@yeah.net; Tel: 022-23909448, E-mail: shilihongxys@ihcams.ac.cn
Received: January 13, 2016 Accepted: May 6, 2016
This work was supported by the National Natural Science Foundation of China (Grant No.31471291)
*Corresponding authors. Tel: +86-22-23909446, E-mail:liufeng198308@yeah.net; Tel: +86-22-23909448, E-mail: shilihongxys@ihcams.ac.cn
网络出版时间: 2016-08-10 11:08:56 URL: http://www.cnki.net/kcms/detail/31.2035.Q.20160810.1108.004.html

1　LncRNA的发现与鉴定

1.1　LncRNA的发现

LncRNA的发现主要依赖于转录组学高通量测序技术。现有的测序技术有微阵列技术(microarrays)、第二代高通量转录组测序技术(second-generation RNA-sequencing technology, RNA-Seq)、单细胞测序技术(single-cell sequencing)、单分子测序技术(single molecule sequencing, SMS)、RNA捕获测序技术(RNA capture-Seq)等。我们在此着重介绍发现lncRNAs的三个主流技术: RNA-Seq、单细胞测序技术以及单分子测序技术(single molecule sequencing, SMS)。

1.1.1　*利用RNA-Seq发现lncRNA*　RNA-Seq是当前转录组测序的主流技术。它不需要预制探针, 而是将测序文库(单链cDNA)连接到测序载体上, 在合成互补链时通过捕捉不同核苷酸的荧光信号来得到测序结果。该技术能检测任何物种转录组的表达, 具有实验成本低、背景信号低、转录子表达丰度测量准确、重复性好等优点。在lncRNA研究中, 绝大多数lncRNA的发现都得益于该技术[3,6]。不过, RNA-Seq技术对高丰度转录本有一定的偏好性, 而由于lncRNA的表达水平普遍偏低, 利用它来建立lncRNA的表达谱时需要适当增加测序深度[7]。

1.1.2　*利用单细胞测序技术发现lncRNA*　RNA-Seq技术对待检测样本量有一定的要求, 无法针对少量、珍贵样本进行转录组测序。单细胞测序可将检测样本量降至单细胞, 对其深度测序可获得这些少量、珍贵样本的转录表达谱(包括lncRNA), 同时该技术还适用于研究异质性细胞群体中单细胞lncRNA的表达谱。Kim等[8]首次将该技术用于lncRNA的研究, 发现在诱导体细胞重编程为多能干细胞(induced pluripotent stem cell, iPSC)过程中lncRNA表达呈动态变化, 重编程过程中被激活的lncRNA能够抑制谱系分化特异基因表达, 促使细胞获得全能性, 提示lncRNA在细胞命运转换中起重要作用。该研究首次证明了利用单细胞测序技术研究lncRNA的可行性, 为将来在组织干细胞、肿瘤干细胞等单细胞水平研究lncRNA提供了参考。不过, 该技术也有局限性, 例如测序结果受RNA逆转录酶的效率、PCR扩增误差等的影响[9-10], 解读测序结果时需要考虑这些因素。

1.1.3　*单分子测序技术有望更加准确地检测lncRNA的表达*　在上述测序技术的基础上提出了第三代测序技术(third-generation sequencing, TGS), 即单分子测序技术, 它代表了测序技术的未来。SMS在测序过程中, 不需要PCR富集扩增, 而是直接将DNA或者cDNA固定在芯片上, 在其互补链合成过程中利用不同荧光标记的dNTP进行测序[11]。该技术避免了PCR扩增带来的系统偏差, 但是仍面临荧光信号弱, 精确捕获难, 测序结果错误率较高等问题, 尚处于开发阶段, 还没有应用于lncRNA的研究。在不久的将来, 随着灵敏度更高的信号识别系统的开发, 该技术将更加成熟完善[12], 届时将为lncRNA研究提供更加全面准确的信息。

1.2　LncRNA的鉴定

1.2.1　*鉴定转录子蛋白质编码潜能*　鉴定转录子蛋白质编码潜能的方法有软件预测和实验鉴定两种[13]。目前预测转录本编码能力的常用软件有BLASTX[14]、CPAT[15]、PhyloCSF[16]、CPC[17]、CNCI[18]等(表1)。Luo等[19]使用CPAT软件分析造血干细胞(hematopoietic stem cells, HSCs)中新发现转录子的编码潜能, 建立了HSCs中lncRNA的表达谱, 并首次研究发现, lncRNA lncHSC-1和lncHSC-2调控HSCs的自我更新和分化。由于软件各自的局限性, 联合使用多个软件可以更加全面地对转录子的编码潜能进行预测。例如, Wang等[20]首先利用PhyloCSF软件预测转录本的编码潜力, 随后结合BLASTX软件从同源序列保守方面进一步筛查, 最终发现了1 000多个大鼠组织中特异表达的lncRNA。然而, 这些常用软件需要对核苷酸进行序列比对, 运行时间较长, 并且受已有数据的限制。为了解决以上难题, 新开发的软件在算法上进行了改进, 引入了更多的参数(例如转录子结构、表达水平等)并加快了运行速度(PLEK)[21], 从而能快速准确地预测转录子的编码潜能。

除软件外, 还可以通过实验手段鉴定lncRNA的编码潜能。LncRNA大部分位于细胞核内, 分布于细胞质内的lncRNA可结合在核糖体上或自由分布于细胞质中。通过核糖体图谱(ribosome profiling assay or Ribo-seq)能检测转录子是否结合在核糖体上, 从而初步判断其是否具有蛋白质编码潜能。但是, 转录子即使结合在核糖体上, 也并不代表有蛋白质编码能力, 它们很可能是利用核糖体进行迁移。研究发现, 在翻译过程中核糖体遇到mRNA终止子

(stop codon)时将使mRNA快速脱离, 而随着迁移的ncRNA则不然。所以结合核糖体图谱和转录子在终止子位置从核糖体上脱离的速度(ribosome release score, RRS)可以鉴别ncRNA[22]。核糖体图谱适用于在组学水平检测转录子的编码潜能, 而针对特定的转录子, 可以利用末端标记法(terminal tag assay)来甄别其编码潜能, 例如lncDC[23]鉴定。

1.2.2 鉴定lncRNA的表达 Northern blot和实时荧光定量PCR(RT-qPCR)是当前实验检测lncRNA表达的主流技术。Northern blot是基于碱基互补配对原理针对目标基因设计带有放射标记的探针, 探针与目的片段杂交放射显影后检测lncRNA的表达。该技术的优点在于能准确检测lncRNA表达丰度及转录子长度, 但是由于探针需要进行放射性标记, 有潜在的安全隐患以及放射污染问题。RT-qPCR是将RNA逆转录为cDNA, 在PCR扩增过程中利用体系中的荧光信号实时定量监测cDNA含量的变化。RT-qPCR检测相对快捷、方便, 已经逐渐成为检测lncRNA表达的主流技术。但是需要注意RT-qPCR的结果依赖于逆转录酶的稳定性, 此外, RT-qPCR的数据分析主观性较强[24]。为了克服各个方法的缺点, 有的研究同时利用两种方法来研究lncRNA的表达, 例如lnc-DC[23]、lncRNA Haunt[19]等。

2 LncRNA的结构和功能检测

2.1 LncRNA的结构分析方法

LncRNA的一级核苷酸序列在进化中缺乏保守性, 其功能的实现主要依赖于二级乃至高级结构[25]。RNA的二级结构可以通过软件预测和实验检测来获悉。预测lncRNA二级结构常用的软件有RNAstructure[26]、RNAfold[27]、Mfold[28]、CentroidFold[29]、UNAfold[30]等。Khaitan等[31]用RNAstructure和RNAfold预测发现, lncRNA SPRY4-IT1的二级结构中有长而稳定的发夹结构。Kino等[32]用UNAFold发现, lncRNAs Gas5和糖皮质激素受体相互作用的区域有6个二级发夹结构。这些研究都为其机制研究提供了参考。

RNA的二级结构除了用软件预测外还可以通过实验方法获得, 如FragSeq[33]、SHAPE-seq[34]、DMS-seq[35]、Mod-seq[36]、MAP-seq[37]和PARS[38]等方法。这些方法各有优劣, 有的利用核酸酶, 有的利用探针来检测lncRNA的结构。

2.2 LncRNA的功能检测

2.2.1 LncRNA敲减、过表达、CRISPR/Cas技术 与研究蛋白质编码基因的方法相似, lncRNA的功能研究也主要采用敲减、过表达、敲除等实验手段。LncRNA的敲减是利用RNA干扰(RNA interference, RNAi)方法, 即由小干扰RNA(small interfering RNAs, siRNAs)和短发夹RNA(short hairpin RNAs, shRNAs)介导的特异性降解同源RNA, 该方法是lncRNA功能研究中最主要、最常见的技术之一。例如l, ncRNA Rmrp[39]、LINC00657[40]、lncRNA Evx1[41]、lnc-TCF7[42]等功能的阐明都用此方法。由于绝大多数lncRNA的表达丰度偏低, 在组织细胞中过表达

表1 常用鉴定基因编码潜能的软件

Table 1 Key tools used for prediction of coding potential of the transcripts

软件名称 Name of software	软件原理 Features and methodology used
BLASTX	BLASTX is based on the conservation of open reading frame, through comparing the similarity of homologous series code region to distinguish protein-coding and non-coding sequences[14].
Codingpotential assessment tool (CPAT)	CPAT uses a logisticregression model built with four sequence features: open reading frame size, open reading framecoverage, Fickett TESTCODE statistic and hexamerusage bias[15].
PhyloCSF	PhyloCSF methodfully leverages multiple alignments in a phylogenetic framework, produces meaningful likelihood ratios as its output and rests upon thesweeping theoretical foundation for statistical model comparison[16].
Coding potential calculator (CPC)	Coding potential calculator (CPC), to assess the protein-coding potential of a transcript based on six biologically meaningful sequence features[17].
CNCI	Coding-non coding index (CNCI), by profiling adjoining nucleotide triplets to effectively distinguish protein-codingand non-coding sequences independent of known annotations[18].

lncRNA也是研究其功能的一个有效策略, 过表达在细胞水平可以利用转染或者病毒介导的感染来实现(例如lncRNA ZEB1-AS1[43]、lncRNA MEG3[44]、lnc-DC[23]等的功能), 而在动物水平则通过建立转基因动物模型来研究其功能。

相比其他基因编辑技术, CRISPR/Cas系统具有省时、高效、精准等优点, 正在成为lncRNA功能研究的主流技术。CRISPR/Cas9技术通过在基因组特定位点上进行切割, 产生双链DNA断点, 诱发细胞进行非同源末端连接(non-homologous end joining, NHEJ)修复DNA损伤。利用CRISPR/Cas9技术敲除lncRNA时, 需要在lncRNAs的两端同时产生DNA双链的断裂, 从而达到敲除的目的。2015年, Yin等[45]用这一技术在小鼠胚胎干细胞中首次实现了lncRNA Haunt基因大片段高效敲除(58 Kb)以及功能区域(例如启动子区、部分外显子区等)定点敲入(knockin)。得益于CRISPR/Cas9技术, 该研究首次揭示了lncRNA存在一种全新调控机制, 即lncRNA的基因组DNA序列和转录子具有不同的功能。随后, Chen等[46]在研究技术上将CRISPR/Cas9和Flp/FRT、Cre/LoxP系统相结合, 在人胚胎干细胞中首次建立了可诱导的基因敲除体系(conditional knockout system, CKO), 为将来在人早期发育过程中条件性敲除lncRNA奠定了基础。此外, 该技术还具有同步敲除多个基因的优势, 为研究功能冗余的lncRNA提供了便利。CRISPR/Cas系统敲除技术优势明显, 但是它潜在的脱靶效应仍值得注意[47]。

2.2.2　RNA荧光原位杂交(RNA-FISH)　RNA荧光原位杂交(RNA-FISH)技术是检测lncRNA亚细胞定位的主流技术, 具有简便、灵敏等特点。LncRNA的作用方式会依据其亚细胞定位的不同而有所差异: 核内的lncRNA多与增强子调控相关(enhancer lncRNA), 而细胞质中的lncRNA多进行转录后调控[19]。RNA-FISH技术针对靶RNA合成带有荧光标记的寡核苷酸探针, 在荧光显微镜观察下根据荧光位置确定靶基因的细胞定位。利用该技术已经成功定位了lncHSC-1[19]、lncHSC-2[19]、Tsix[48]、Xist[49]和lncTCF7[42]等多个lncRNAs。

2.2.3　c-KLAN技术　非编码RNA沉默与定位分析(combined knockdown and localization analysis of non-codingRNAs, c-KLAN)主要用于对lncRNA进行功能缺失和细胞定位研究。该技术利用靶lncRNA的cDNA序列体外转录生成dsRNA, 用核糖核酸内切酶III酶切制备esiRNA, esiRNA混合物可在多个位点与靶RNA结合, 从而达到高效抑制靶基因表达的目的。与其他RNAi方法相比, esiRNA明显降低了脱靶效应, 实验结果更加可靠。同时, 扩增cDNA还可以用来制备FISH探针, 来对细胞或组织中的lncRNA进行定位分析。Chakraborty等[50]提出并采用c-KLAN技术确定了Panctl1主要分布于胚胎干细胞的细胞核中, 进一步研究证实Panctl1对于胚胎干细胞多能性的维持起着决定性作用。

3　LncRNA生物学机制研究

LncRNAs能利用其一级核苷酸序列或高级结构来结合DNA、RNA以及蛋白质来发挥多种功能。以下我们从检测与lncRNAs相互作用的分子以及lncRNA和靶基因结合的空间构象两大方面, 阐述lncRNAs发生生物学过程中的调控机制。

3.1　LncRNA-DNA、RNA、蛋白质相互作用研究方法

3.1.1　ChIRP技术　ChIRP(chromatin isolation by RNA purification)通过设计生物素或链霉亲和素探针来钓取靶RNA, 将与其相互作用的DNA、RNA片段或者相互作用蛋白富集, 该技术与高通量测序技术(DNA-seq、RNA-seq)以及质谱(mass spectrum)技术结合能在组学水平分别揭示与lncRNAs所结合的DNA、RNA和蛋白[51], 而与qPCR、免疫印迹法等方法结合能在个体水平检测相互作用的分子。例如, 将lncHSC-2 ChIRP后进行DNA-seq揭示了其在基因组上结合位点[19]。Chu等[51]运用ChIRP-seq技术确定lncRNA HOTAIR在基因组上的定位(优先定位于富含GA的DNA区域)。

最近, Quinn等[52]对ChIRP进行改进, 提出了dChIRP(domain-specific chromatin isolation by RNA purification)新技术。该技术在天然环境下剖析了lncRNAs不同结构域的功能, 为lncRNAs功能的研究提供了实用性的工具。dChIRP技术设计生物素化的反义寡核苷酸池来靶标lncRNAs的特定结构域, 并采用与ChIRP相同的方法分离纯化与lncRNAs相互作用的复合物。Quinn等[52]成功运用该技术研究了果蝇X染色质剂量补偿调控基因lncRNA roX1, 发现roX1的功能域为“三指手掌”结构。

3.1.2　RNA免疫沉淀(RNA-immuno precipitation,

RIP) RNA免疫沉淀可以用来研究组织或活细胞中蛋白质和RNA之间的相互作用, 是研究lncRNA的常用方法之一。该方法的原理是采用针对目标蛋白的抗体, 把相应的RNA-蛋白复合物沉淀下来, 经过分离纯化, 并对复合物上的RNA进行分析鉴定[53]。Liu等[54]用这一技术证实了lncRNA GAS5和YBX1蛋白间存在相互作用。Pandey等[55]用RIP技术揭示, Kcnq1ot1与H3K9甲基转移酶和PRC2复合物之间存在相互作用。另外, RIP技术分别与Chip和RNA-Seq相结合而衍生的RIP-Chip和RIP-Seq技术, 也广泛应用于lncRNA与蛋白质的相互作用中。Sheik等[56]用RIP-Chip技术揭示了一些保守的lncRNAs的转录是由Nanog以及Pou5f1调控的。

3.1.3 紫外交联免疫沉淀法测序(CLIP seq) 紫外交联免疫沉淀法(cross linking and immune precipitation, CLIP)是一项在全基因组水平上揭示lncRNA与RNA结合蛋白相互作用的革命性技术。其主要原理是基于RNA分子与RNA结合蛋白在紫外照射下发生偶联, 以RNA结合蛋白的特异性抗体将lncRNA-蛋白质复合体沉淀, 收集RNA片段并进行高通量深入测序, 与lncRNA结合的蛋白质可通过液相色谱–质谱联用技术(HPLC-MS)或SDS-PAGE分析鉴定。该技术不仅提高了RNA与蛋白质的特异性结合, 还可以识别出它们之间的结合位点[57]。利用CLIP技术已经检测到了多种结合在PRC2蛋白复合物上的lncRNA, 例如ANRIL(hallmark)、Air等。

目前, 以CLIP技术为基础衍生出了单核苷酸分辨率交联–免疫共沉淀法(iCLIP)、光活性增强的核糖核苷交联和免疫共沉淀法(PAR-CLIP)。iCLIP根据交联位点位于cDNA截断位置上游一个核苷酸的位置, 将RNA与相互作用蛋白质结合位点的分辨率提高至单核苷酸水平[58], 例如hnRNPL蛋白质结合的RNA[59]。 PAR-CLIP是根据紫外光线能够增加含有4-硫尿核苷(4-thiouridine, 4SU)的RNA和蛋白质的交联效率, 而处于RNA和蛋白质结合位点的4SU, 在逆转录为cDNA时会发生T-C的突变, 所以深度测序突变位点就能准确提供RNA与蛋白质的结合位点[60]。例如Yoon等[61]通过PAR-CLIP技术发现了和AUF1(AU-binding factor 1, 一种RNA结合蛋白)结合的RNA富含U-/GU-序列。然而, iCLIP和RAP-CLIP技术在检测蛋白所结合RNA时都需要进行PCR扩增, 无法避免PCR的碱基偏好性, 所以仍需要继续改进方法以提高分辨率。

3.2 LncRNA和靶基因的空间构象

了解lncRNA与其靶基因的空间构象有助于阐明lncRNA的调控机制。染色体构象捕获技术

表2 常用lncRNA数据库简介

Table 2 Key databases of lncRNA

数据库名称及网址 Database and web link	数据库功能 Description
LncRNAdb (http://www.lncRNAdb.org/)	It contains lots of information about the lncRNAs including sequences, structural information, genomic context, expression, subcellular localization, conservation, functional evidence and other relevant information[65].
NONCODE (http://www.bioinfo.org/noncode/)	It includes almost all types of ncRNAs, except transfer RNAs and ribosomal RNAs[66]. NONCODE v3.0 includes the first integrated collection of expression and functional lncRNA data obtained from re-annotated microarray studies. The lncRNAdb and NONCODE include a BLAST search tool and easy export of content via direct download[67-68].
ChIPBase (http://deepbase.sysu.edu.cn/chipbase/)	It facilitates the comprehensive annotation and discovery of transcription factor binding maps and transcriptional regulatory relationships of lncRNAs and miRNAs from ChIP-Seq data[69].
LncRNADisease (http://cmbi.bjmu.edu.cn/lncrnadisease)	It reveals lncRNAs' roles in diseases and identify candidate lncRNAs for disease diagnosis, treatment and prognosis[70].
NPInter (http://www.bioinfo.org/NPInter)	It integrates experimentally verified functional interactions between noncoding RNAs (excluding tRNAs and rRNAs) and other biomolecules (proteins, RNAs and genomic DNAs)[71].
StarBase (http://starbase.sysu.edu.cn/)	It provides the most comprehensive CLIP-Seq experimentally supported miRNA-mRNA and miRNA-lncRNA interaction networks to date[72-73].
LncRNome (http://genome.igib.res.in/lncRNome)	It is one of the largest and most comprehensive resources for lncRNAs, including structures, dbSNP, interactions with proteins and other RNAs[74].
LNCipedia (http://www.lncipedia.org)	LNCipedia offers 21488 annotated human lncRNA transcripts obtained from different sources. In addition to basic transcript information and gene structure, several statistics are determined for each entry in the database, such as secondary structure information, protein coding potential and microRNA binding sites[75].

(chromosome conformation capture, 3C)是在目的lncRNA及其下游靶基因核酸序列都已知的前提下，完成“一对一”的空间位置构象研究。Sun等[62]用3C技术证明了lncRNA IRAIN基因的启动子与增强子之间存在相互作用。由染色体构象捕获技术而衍生的环状染色质构象捕获技术(circular chromosome conformation capture, 4C)、3C碳拷贝技术(3C-carbon copy, 5C)技术、Hi-C技术对与lncRNA相互作用的染色体序列分别可以进行“一对多”(4C)与“多对多”(5C和Hi-C)的研究。Zhao等[63]用4C技术研究lncRNA H19相关的基因印记区, 发现H19印记区不仅在染色体的物理结构上起作用, 而且影响着基因的表观遗传状态。Ma等[64]在Hi-C技术基础上, 用Dnase I代替了限制性内切酶, 使得Hi-C的效率和分辨率都得到了很大的提高。他们用这项技术揭示了两个人类细胞系中的998个lncRNAs基因在基因组的空间构象。

3.3 LncRNA数据库

此外, 随着lncRNAs的发现和数据积累, 相关的数据库也日臻完善, 可以利用查阅数据库来检索lncRNA的相关信息(表2)。

4 展望

本文总结了研究lncRNAs的常用方法。得益于上述研究技术和方法的日渐成熟, lncRNAs的研究取得了很大进展。不过部分技术方法依然存在一定的局限性, 其突破有望加速推动我们对于lncRNAs的认识。例如, 建立解析体内lncRNAs的二级乃至高级结构技术和方法。现有的方法虽然可以获得lncRNAs的二级及高级结构, 但都是在体外进行, 由于lncRNAs在体外和体内的二级结构可能不同, 如何研究体内lncRNAs的结构, 并在此基础上研究其功能和作用机制是该领域研究面临的挑战之一。在基因编辑方面, 如何利用CRISPR/Cas9基因编辑技术高效地进行lncRNAs大片段敲除, 基因的定点整合也是一项关键的技术突破点。而第三代测序技术和单细胞测序技术的完善将有助于发现新的、特异性表达的lncRNAs。此外, 开发新技术在体内鉴定lncRNAs编码的短肽并研究其功能[76-77]也将是一个重要的研究方向。

参考文献 (References)

1　Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al*. Initial sequencing and analysis of the human genome. Nature 2001; 409(6822): 860-921.

2　Flynn RA, Chang HY. Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell 2014; 14(6): 752-61.

3　Yang L, Froberg JE, Lee JT. Long noncoding RNAs: Fresh perspectives into the RNA world. Trends Biochem Sci 2014; 39(1): 35-43.

4　Philippen LE, Dirkx E, da Costa-Martins PA, de Windt LJ. Non-coding RNA in control of gene regulatory programs in cardiac development and disease. J Mol Cell Cardiol 2015; 89(Pt A): 51-8.

5　Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: Evidence of widespread transcription detected by microarray tiling experiments. Trends Genet 2005; 21(2): 93-102. .

6　Ilott NE, Ponting CP. Predicting long non-coding RNAs using RNA sequencing. Methods 2013; 63(1): 50-9.

7　Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell 2009; 136(4): 629-41.

8　Kim DH, Marinov GK, Pepke S, Singer ZS, He P, Williams B, *et al*. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. Cell Stem Cell 2015; 16(1): 88-101.

9　Sandberg R. Entering the era of single-cell transcriptomics in biology and medicine. Nat Methods 2014; 11(1): 22-4.

10　Eberwine J, Sul JY, Bartfai T, Kim J. The promise of single-cell sequencing. Nat Methods 2014; 11(1): 25-7.

11　Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, *et al*. Single-molecule DNA sequencing of a viral genome. Science 2008; 320(5872): 106-9.

12　Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010; 19(R2): R227-40.

13　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. Biochim Biophys Acta 2016; 1859(1): 31-40.

14　Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet 1993; 3(3): 266-72.

15　Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res 2013; 41(6): e74.

16　Lin MF, Jungreis I, Kellis M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 2011; 27(13): i275-82.

17　Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, *et al*. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 2007; 35(Web Server issue): W345-9.

18　Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, *et al*. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 2013; 41(17): e166.

19　Luo M, Jeong M, Sun D, Park HJ, Rodriguez BA, Xia Z, *et al*. Long non-coding RNAs control hematopoietic stem cell function. Cell Stem Cell 2015; 16(4): 426-38.

20　Wang F, Li L, Xu H, Liu Y, Yang C, Cowley AW Jr, *et al*.

Characteristics of long non-coding RNAs in the Brown Norway rat and alterations in the Dahl salt-sensitive rat. Sci Rep 2014; 4: 7146.
21 Li A, Zhang J, Zhou Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics 2014; 15: 311.
22 Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 2013; 154(1): 240-51.
23 Wang P, Xue Y, Han Y, Lin L, Wu C, Xu S, *et al*. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science 2014; 344(6181): 310-3.
24 Nolan T, Hands RE, Bustin SA. Quantification ofmRNA using real-time RT-PCR. Nat Protoc 2006; 1(3): 1559-82.
25 Guo X, Gao L, Wang Y, Chiu DK, Wang T, Deng Y. Advances in long noncoding RNAs: Identification, structure prediction and function annotation. Brief Funct Genomics 2016; 15(1): 38-46.
26 Mathews DH. Using the RNAstructure software package to predict conserved RNA structures. Curr Protoc Bioinformatics 2014; 46: 12.4.1-22.
27 Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, *et al*. ViennaRNA Package 2.0. Algorithms Mol Biol 2011; doi: 10.1186/1748-7188-6-26.
28 Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31(13): 3406-15.
29 Sato K, Hamada M, Asai K, Mituyama T. CENTROIDFOLD: A web server for RNA secondary structure prediction. Nucleic Acids Res 2009; 37(Web Server issue): W277-80.
30 Markham NR, Zuker M. UNAFold: Software for nucleic acid folding and hybridization. Methods Mol Biol 2008; 453: 3-31.
31 Khaitan D, Dinger ME, Mazar J, Crawford J, Smith MA, Mattick JS, *et al*. The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. Cancer Res 2011; 71(11): 3852-62.
32 Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 2010; 3(107): ra8.
33 Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, *et al*. FragSeq: Transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods 2010; 7(12): 995-1001.
34 Loughrey D, Watters KE, Settle AH, Lucks JB. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res 2014; doi: 10.1093/nar/gku909.
35 Lightfoot HL, Hall J. Endogenous polyamine function–the RNA perspective. Nucleic Acids Res 2014; 42(18): 11275-90.
36 Lin Y, May GE, Joel McManus C. Mod-seq: A high-throughput method for probing RNA secondary structure. Methods Enzymol 2015; 558: 125-52.
37 Seetin MG, Kladwang W, Bida JP, Das R. Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. Methods Mol Biol 2014; 1086: 95-117.
38 Vinogradova SV, Sutormin RA, Mironov AA, Soldatov RA. Probing-directed identification of novel structured RNAs. RNA biology 2016; 13(2): 232-42.
39 Huang W, Thomas B, Flynn RA, Gavzy SJ, Wu L, Kim SV, *et al*. DDX5 and its associated lncRNA Rmrp modulate TH17 cell effector functions. Nature 2015; 528(7583): 517-22.
40 Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, *et al*. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. Cell 2016; 164(1/2): 69-80.
41 Luo S, Lu JY, Liu L, Yin Y, Chen C, Han X, *et al*. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. Cell Stem Cell 2016; 18(5): 637-52..
42 Wang Y, He L, Du Y, Zhu P, Huang G, Luo J, *et al*. The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. Cell Stem Cell 2015; 16(4): 413-25.
43 Li T, Xie J, Shen C, Cheng D, Shi Y, Wu Z, *et al*. Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma. Oncogene 2016; 35(12): 1575-84.
44 Chang L, Wang G, Jia T, Zhang L, Li Y, Han Y, *et al*. Armored long non-coding RNA MEG3 targeting EGFR based on recombinant MS2 bacteriophage virus-like particles against hepatocellular carcinoma. Oncotarget 2016; 7(17): 23988-4004.
45 Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, *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(5): 504-16.
46 Chen Y, Cao J, Xiong M, Petersen AJ, Dong Y, Tao Y, *et al*. Engineering human stem cell lines with inducible gene knockout using CRISPR/Cas9. Cell Stem Cell 2015; 17(2): 233-44.
47 Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 2013; 31(9): 839-43.
48 Maclary E, Buttigieg E, Hinten M, Gayen S, Harris C, Sarkar MK, *et al*. Differentiation-dependent requirement of Tsix long non-coding RNA in imprinted X-chromosome inactivation. Nat Commun 2014; 5: 4209.
49 McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, *et al*. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 2015; 521(7551): 232-6.
50 Chakraborty D, Kappei D, Theis M, Nitzsche A, Ding L, Paszkowski-Rogacz M, *et al*. Combined RNAi and localization for functionally dissecting long noncoding RNAs. Nat Methods 2012; 9(4): 360-2.
51 Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 2011; 44(4): 667-78.
52 Quinn JJ, Ilik IA, 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. Nat Biotechnol 2014; 32(9): 933-40.
53 Tenenbaum SA, Carson CC, Lager PJ, Keene JD. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci USA 2000; 97(26): 14085-90.
54 Liu Y, Zhao J, Zhang W, Gan J, Hu C, Huang G, *et al*. lncRNA GAS5 enhances G_1 cell cycle arrest via binding to YBX1 to

regulate p21 expression in stomach cancer. Sci Rep 2015; 5: 10159.

55 Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, *et al*. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 2008; 32(2): 232-46.

56 Sheik Mohamed J, Gaughwin PM, Lim B, Robson P, Lipovich L. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA 2010; 16(2): 324-37.

57 Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. Science 2003; 302(5648): 1212-5.

58 Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, *et al*. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 2010; 17(7): 909-15.

59 Ruan X, Li P, Cangelosi A, Yang L, Cao H. A long non-coding RNA, lncLGR, regulates hepatic glucokinase expression and glycogen storage during fasting. Cell Rep 2016; 14(8): 1867-75.

60 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, *et al*. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010; 141(1): 129-41.

61 Yoon JH, De S, Srikantan S, Abdelmohsen K, Grammatikakis I, Kim J, *et al*. PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. Nat Commun 2014; 5: 5248.

62 Sun J, Li W, Sun Y, Yu D, Wen X, Wang H, *et al*. A novel antisense long noncoding RNA within the IGF1R gene locus is imprinted in hematopoietic malignancies. Nucleic Acids Res 2014; 42(15): 9588-601.

63 Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, *et al*. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 2006; 38(11): 1341-7.

64 Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, *et al*. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat Methods 2015; 12(1): 71-8.

65 Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: A reference database for long noncoding RNAs. Nucleic Acids Res 2011; 39(Database issue): D146-51.

66 Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, *et al*. NONCODE: An integrated knowledge database of non-coding RNAs. Nucleic Acids Res 2005; 33(Database issue): D112-5.

67 Bu D, Yu K, Sun S, Xie C, Skogerbo G, Miao R, *et al*. NONCODE v3.0: Integrative annotation of long noncoding RNAs. Nucleic Acids Res 2012; 40(Database issue): D210-5.

68 Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, *et al*. lncRNAdb v2.0: Expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res 2015; 43(Database issue): D168-73.

69 Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: A database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res 2013; 41(Database issue): D177-87.

70 Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, *et al*. LncRNA disease: A database for long-non-coding RNA-associated diseases. Nucleic Acids Res 2013; 41(Database issue): D983-6.

71 Yuan J, Wu W, Xie C, Zhao G, Zhao Y, Chen R. NPInter v2.0: An updated database of ncRNA interactions. Nucleic Acids Res 2014; 42(Database issue): D104-8.

72 Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res 2011; 39(Database issue): D202-9.

73 Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 2014; 42(Database issue): D92-7.

74 Bhartiya D, Pal K, Ghosh S, Kapoor S, Jalali S, Panwar B, *et al*. lncRNome: A comprehensive knowledgebase of human long noncoding RNAs. Database (Oxford) 2013; 2013: bat034.

75 Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, *et al*. LNCipedia: A database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013; 41(Database issue): D246-51.

76 Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, *et al*. Toddler: An embryonic signal that promotes cell movement via Apelin receptors. Science 2014; 343(6172): 1248636.

77 Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, *et al*. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 2015; 160(4): 595-606.