

doi: 10.3969/j.issn.0490-6756.2017.03.038

# 一种改进的小非编码 RNA cDNA 文库构建方法

杜光石, 罗发涛, 肖燕, 陈红军, 吴传芳

(四川大学生命科学学院功能基因组实验室, 成都 610064)

**摘要:** 本文使用 RNA Antisense Purification (RAP) 技术富集纯化 sncRNA, 有效去除了 5S、5.8S rRNA. 联合使用 Tobacco Acid Pyrophosphatase (TAP) 和 RNA 5' Polyphosphatase 处理 sncRNA, 将 5'-端为非单磷酸结构的 sncRNA 转变成单磷酸结构, 增加了文库覆盖率. 在 sncRNA 3'-端添加 poly(A) 尾结构以及使用 oligo(dT)<sub>16</sub> VN-adaptor 锚定引物进行逆转录, 间接解决了 sncRNA 3'-端接头定向连接问题, 不再使用价格昂贵的腺苷化接头. 改进后的 sncRNA cDNA 文库容量为  $3 \times 10^5$  cfu, 重组率为 95%, 且测序结果显示, 获得的 cDNA 序列具有较好的完整性. 使用这种成本低且行之有效的改进方法, 为深入研究 sncRNA 奠定了基础.

**关键词:** sncRNA; cDNA 文库; poly(A); RNA 反义链纯化 (RAP)

**中图分类号:** Q78 **文献标识码:** A **文章编号:** 0490-6756(2017)03-0653-05

## An improved method of sncRNA cDNA library construction

DU Guang-Shi, LUO Fa-Tao, XIAO Yan, CHEN Hong-Jun, WU Chuan-Fan

(Department of Function Genome, College of Life Science, Sichuan University, Chengdu 610064, China)

**Abstract:** 5S and 5.8S rRNA were removed efficiently by RNA Antisense Purification (RAP). More 5'-monophosphorylated RNAs were acquired after using Tobacco Acid Pyrophosphatase (TAP) and RNA 5' sncRNA Polyphosphatase, which resulted in the increasement of the coverage of the cDNA library. To avoid using the adenylated adaptor, a poly(A) structure was added to sncRNA 3' end and then reverse transcription was carried out with oligo(dT)<sub>16</sub> VN-adaptor primers. This improved sncRNA cDNA library contained  $3 \times 10^5$  independent clones and the recombination rate was 95%. The sequencing results showed that the sncRNA cDNA sequences had good integrity by employing this improved and cost-effective method, which laid the foundation for the further study of sncRNA.

**Keywords:** sncRNA; cDNA library; poly(A); RAP

## 1 引言

人类基因组中, 70% 左右的序列能够转录为 RNA<sup>[1,2]</sup>, 但多数不能被翻译成蛋白, 这些非编码 RNA (noncoding RNA, ncRNA) 根据长度分为长非编码 RNA (>200 nt, long noncoding RNA, lncRNA) 和小非编码 RNA (<200 nt, small non-

coding RNA, sncRNA)<sup>[3]</sup>.

研究表明<sup>[4]</sup>, sncRNA 在人类基因表达调控、细胞生长、分化、增殖以及疾病的发生等方面发挥重要作用, 其中又以 microRNA (miRNA) 和 Piwi-associated RNA (piRNA) 研究得最深入. miRNA 由较长的初级转录物 (pre-miRNA) 经过 Drosha 与 Dicer 等核酸酶加工、剪切后, 与其它蛋白组装成

收稿日期: 2015-03-05

基金项目: 国家自然科学基金项目 (31371325)

作者简介: 杜光石 (1990-), 男, 贵州贵阳人, 硕士研究生, 研究方向为细胞的分子生物学. E-mail: duguangshi2013@163.com

通讯作者: 吴传芳. E-mail: wuchuanfang@scu.edu.cn

RNA 诱导沉默复合体 (RNA-induced silencing complex, RISC), 再以序列配对的方式与靶 mRNA 的 3'-端非翻译区 (3' UTR) 特异结合, 引起靶 mRNA 降解或翻译抑制实现对基因表达的调控<sup>[5-7]</sup>. piRNA 最早在哺乳动物睾丸组织中发现<sup>[8]</sup>, 其在基因沉默和表观遗传调控等方面发挥着重要作用<sup>[9,10]</sup>. 此外, 对 small nuclear RNA (snRNA)<sup>[11]</sup>, transfer RNA (tRNA)<sup>[12]</sup>, small interfering RNA (siRNA)<sup>[13]</sup> 与 small nucleolar RNA (snoRNA)<sup>[14]</sup> 等多种 sncRNA 的研究, 也进一步揭示了它们在生命各个层次中发挥的重要功能.

目前, 我们对 sncRNA 的认识还十分有限, 这也意味着对 sncRNA 研究有很大的探索空间, 而构建 sncRNA cDNA 文库是发现和研究 sncRNA 的主要途径. 多数 sncRNA 不带 poly(A) 尾结构<sup>[15,16]</sup>, 传统 sncRNA cDNA 文库构建试剂盒的原理基于直接在 sncRNA 3' 与 5'-端各自加上接头 (3'-接头, 5'-接头), 逆转录后再用接头序列引物进行 PCR 扩增 cDNA. 但是, 商品化的试剂盒价格十分昂贵, 这主要是因为用来做定向连接的 3'-接头是修饰性接头, 需要对其进行 5'-端腺苷化修饰 (5', 5'-adenyl pyrophosphoryl moiety), 才能实现接头的末端定向连接<sup>[17]</sup>. 这种修饰性接头不仅成本高, 而且多数生物技术服务公司还不具备合成这种修饰性接头的能力<sup>[18]</sup>. 此外, 一些 sncRNA 的 5'-端并非单磷酸结构, 如 U1、U2、U4、U5 snRNA 等具有 5'-端帽子结构<sup>[19]</sup>, 次级 siRNA 5'-端为三磷酸或二磷酸基团结构<sup>[20]</sup>, 而 T4 RNA 连接酶不能有效连接这类 sncRNA, 这无疑影响了 sncRNA cDNA 文库的覆盖率. 同时, sncRNA 中, 5S、5.8S rRNA 丰度较高<sup>[21]</sup>, 若不提前除去, 必定对后续 sncRNA cDNA 文库筛选工作造成较大干扰, 而市面上 rRNA 去除试剂盒价格也十分昂贵.

考虑到以上问题, 本研究对传统 sncRNA cDNA 文库构建方法进行了改进, 设计了一种成本较低、且行之有效的 sncRNA cDNA 文库构建方法. 使用这种改进的方法, 最终可以获得完整的 sncRNA cDNA 序列, 为深入研究 sncRNA 奠定了基础.

## 2 材料和方法

### 2.1 材料

2.1.1 菌株和细胞株 大肠杆菌 *E. coli* (*Escherichia coli*) 菌株 DH5 $\alpha$  由本实验室保存, 用于重组子的筛选和质粒的扩增. 人胚肾上皮细胞

HEK293 购于美国 ATCC (American Type Culture Collection), 本实验保存.

2.1.2 主要试剂 细胞培养基 DMEM、胎牛血清 (FBS)、DNase I、RevertAid Reverse Transcriptase、dNTP、RNase Inhibitor、Dynabeads<sup>®</sup> MyOne<sup>™</sup> Streptavidin Trial Kit 购于 Thermo 公司 (上海), T4 RNA Ligase、*E. coli* Poly(A) Polymerase 购于 NEB 公司 (北京), Tobacco Acid Pyrophosphatase、RNA 5' Polyphosphatase 购于 Epicentre 公司 (上海), pMD<sup>™</sup> 19-T Vector 购于 TaKaRa 公司 (大连), 小 RNA 提取试剂盒、PCR 产物纯化试剂盒、2 $\times$ PCR mix、2 $\times$ q-PCR mix 购于 Foregene 公司 (成都), 生物素标记的 5S、5.8S DNA 探针合成于金唯智公司 (苏州), 其它化学试剂均为国产分析纯.

表 1 PCR 引物序列

Tab. 1 Sequences of oligonucleotide primers used for PCR

引物	序列 (5'→3')
5S-F	GTCTACGGCCATACCACCCTGAA
5S-R	AAGCCTACAGCACCCGGTATTCC
5.8S-F	TAGCGGTGGATCACTCGGCTC
5.8S-R	CGACGCTCAGACAGGCGTAGC
U1-F	GGGAGATACCATGATCACGAAGGT
U1-R	ATGCAGTCGAGTTTCCCACA
U2-F	TCTCGGCCTTTTGGCTAAGATCA
U2-R	AGTGGACGGAGCAAGCTCCTATT
U6-F	GCTCGCTTCGGCAGCACATATA
U6-R	CGCTTCACGAATTTGCGTGTC
7SK-F	ATCGCCAGGGTTGATTCGGCTGAT
7SK-R	GGATGTGTCTGGAGTCTTGAAGC
pMD19-T-F	CGCCAGGGTTTCCCAGTCACGAC
pMD19-T-R	GAGCGGATAACAATTTACACAGG

(注: F-forward primer, R-reverse primer)

### 2.2 方法

2.2.1 细胞培养 HEK293 细胞在 37 $^{\circ}$ C、5% CO<sub>2</sub> 的细胞培养箱中培养, 培养基为 DMEM 加 10% (v/v) 胎牛血清 (FBS) 和 1% (v/v) 双抗 (青霉素和链霉素). 当细胞生长密度达到 70%~80% 时, 收集细胞.

2.2.2 利用 RAP (RNA Antisense Purification) 富集 sncRNA 用 Foregene 小 RNA 提取试剂盒提取 HEK293 sncRNA, 通过紫外分光光度计测量和琼脂糖凝胶电泳检测其质量. 将 sncRNA 与生物素标记的 5S、5.8S DNA 探针混合, 65 $^{\circ}$ C 变性 5min, 冰浴 2min, 然后置于 37 $^{\circ}$ C 的分子杂交炉中杂交 10min, 使 sncRNA 样品中的 5S、5.8S DNA 探针通过碱基配对与 5S、5.8S rRNA 充分结合.

接下来加入适量链霉亲和素磁珠 M-270,37°C 继续杂交 20min,借助磁力架收集上清,再用 DNase I 处理,降解残余的 DNA 探针,即得到去除 5S、5.8S rRNA 的 snRNA 溶液。实验组和对照组同时平行操作,对照组不加磁珠 M-270,rRNA 的去除效率通过 qPCR 检测。

**2.2.3 cDNA 文库构建** 提取 snRNA 后通过 RAP 技术富集纯化,然后使用 *E. coli* Poly(A) Polymerase 在 snRNA 3' 末端加上 poly(A) 尾,用 Tobacco Acid Pyrophosphatase 和 RNA 5' Polyphosphatase 处理 snRNA,得到 5'-端为单磷酸基团的 snRNA,再用 T4 RNA 连接酶连接 RNA 接头 adapter-1,使用带接头的锚定引物 oligo(dT)<sub>16</sub> VN-adapter-2(V= A/G/C;N= A/G/C/T) 做逆转录,最后利用两端接头特异性引物进行 PCR,将 PCR 产物连入 T 载体,进行 TA 克隆建库。

**2.2.4 cDNA 文库质量检测及容量计算** 使用两端接头引物进行 PCR,琼脂糖凝胶电泳检测产物分布情况,然后将 PCR 产物进行 TA 克隆,计数蓝白斑,测定重组率。随机挑取单克隆,做菌落 PCR 进一步检测 DNA 插入片段情况并提取质粒测序。将测序结果在 NCBI 数据库中进行序列比对分析(<http://www.ncbi.nlm.nih.gov/>)。文库容量使用 Clarke-Carbon 公式计算<sup>[22]</sup>,其中,N=所需克隆数,p=要求的概率(通常为 0.99), $\frac{1}{n}$  = 某一种 RNA 在给定 RNA 中所占的比例。 $N = \ln(1/p) / \ln[1 - (1/n)]$ 。

## 3 实验结果

### 3.1 HEK293 snRNA 提取检测

哺乳动物的核糖体 RNA(28S,18S,5.8S,5S) 在总 RNA 中丰度较高,如果不提前富集 RNA,除去 rRNA,这将会对后续 snRNAcDNA 文库构建造成很大干扰。使用 Foregene 小 RNA 提取试剂盒提取的 snRNA 虽然去除了 28S、18S rRNA 等大片段 RNA(图 1),得到了质量较高的 snRNA,但依然存在大量 5.8S 和 5S 小分子 rRNA。

### 3.2 RAP 富集 snRNA

通过 RAP(RNA antisense purification)方法,进一步去除 5S、5.8S rRNA,qPCR 检测结果显示,5S、5.8S rRNA 去除效率可以达到 95% 以上(图 2),效果显著,进一步富集了 snRNA,提高了

建库起始 snRNA 的质量。

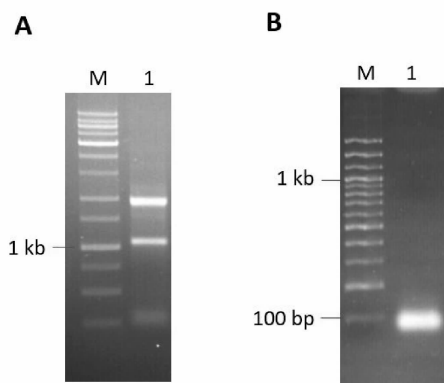


图 1 HEK293 总 RNA 与 snRNA 提取  
A: TRIzol 提取 HEK293 总 RNA。M: marker; 1: HEK293 总 RNA。B: Foregene 小 RNA 提取试剂盒提取 HEK293 snRNA。M: marker; 1: HEK293 snRNA。  
Fig.1 Extraction of HEK293 Total RNA and snRNA  
A: HEK293 Total RNA extraction by TRIzol.  
B: HEK293 snRNA extraction with ForegeneKit.

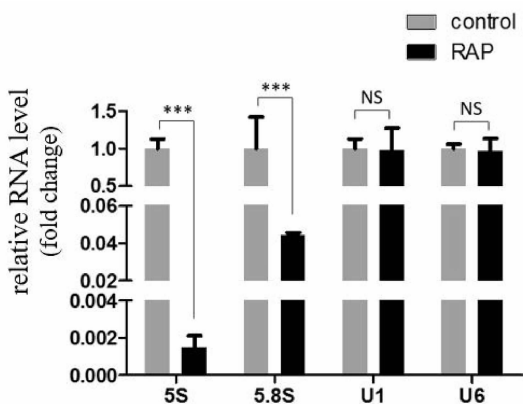


图 2 qPCR 检测 RAP 方法去除 5S、5.8S rRNA 的效率

Fig.2 Quantitative PCR detection of the efficiency to remove 5S and 5.8S rRNA with the method of RNA antisense purification (RAP)

The removal efficiency of 5S and 5.8S rRNA. (Values are shown as mean  $\pm$  SD, \*\*\*  $P < 0.001$ , NS: no significant)

### 3.3 PCR 扩增 cDNA

在 snRNA 5'-端加上接头 adapter-1 后,通过引物 oligo(dT)<sub>16</sub> VN-adapter-2 做逆转录,获得两端带上接头序列的 cDNA 第一链,然后用接头引物 adapter-1/2 进行 PCR,琼脂糖凝胶电泳显示 cDNA 片段分主要分布在 200 bp 以内(图 3),而且具有良好的延展性,说明两端接头添加成功,各个不同长度的 snRNA 片段得到了有效扩增。

### 3.4 snRNA cDNA 文库鉴定

蓝白斑结果显示,文库重组率达到 95% 以上。随机挑取白色菌落,使用载体引物 pMD19-T-F/R

进行菌落 PCR 鉴定分析,结果表明(图 4),PCR 产物大小分布在 200~500bp (若发生自连,PCR 产物大小应为 150bp),说明 cDNA 文库插入频率较高。

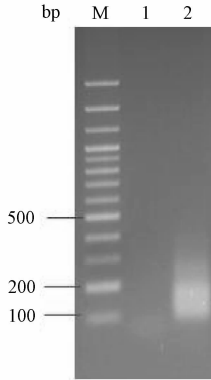


图 3 PCR 扩增 cDNA

M: marker; 1: PCR 负对照,不加 cDNA 模板;  
2: PCR 产物.

Fig. 3 cDNA amplification by PCR

M: marker; 1: PCR without cDNA template;  
2: PCR product.

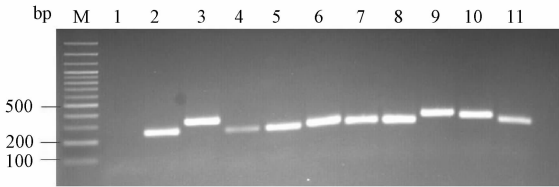


图 4 sncRNA cDNA 文库插入片段的菌落 PCR 鉴定

M: marker; 1: PCR 负对照,不加模板;2-11: 菌落 PCR  
Fig. 4 Identification of insert cDNA by bacterial colony PCR

M: marker; 1: PCR without template; 2-11: bacterial colony PCR

### 3.5 文库容量与测序结果

已构建的 sncRNA cDNA 文库总共包含了约  $3 \times 10^5$  个单克隆重组质粒,根据 Clarke-Carbon 文库计算公式,可以以 99% 概率包含约  $6.5 \times 10^4$  个 sncRNA 的 cDNA 克隆,达到建库要求. 随机挑取 20 个单菌落,提取质粒测序,在 NCBI 数据库中进行序列比对,除去未知序列,发现其中 4 条测序结果与 NCBI 中已知序列的覆盖率达到 99% 以上(表 2),说明 cDNA 具有较好的序列完整性. 虽然本研究测序数量相对较少,接下来还需大量测序结果进一步检测此文库的质量,但从已有测序结果显示,cDNA 具有较好的序列完整性.

## 4 讨论

构建 sncRNA cDNA 文库,对 sncRNA 的研

究具有重要意义. 通常构建这类分子量小的 cDNA 文库会遇到两个问题,一个是如何提取目的 RNA 并富集纯化. 使用 Foregene 公司小 RNA 提取试剂盒,有效的去除了 28S、18S 等大片段 RNA. 再使用 RAP 技术<sup>[23]</sup>,设计与 5.8S、5S rRNA 特异性互补的 DNA 探针,5S、5.8S rRNA 去除率可达 95% 以上,进一步纯化富集了 sncRNA,从而得到起始质量更高的 sncRNA. RAP 实验效果显著,也为今后 RNA 的纯化分离研究提供了一个很好的思路.

表 2 NCBI 序列比对结果

Tab. 2 Sequence alignment in NCBI

RNA	公司测序长度(bp)	NCBI 中长度(bp)	序列完整度(%)
tRNaseZL-interacting RNA A9	39	39	100
Human Val-tRNA-1a	75	76	99
U5B small nuclear 1(RNU5B-1), small nuclear RNA	117	117	100
U2 small nuclear 1(RNU2-1), small nuclear RNA	187	187	100

另一个问题是最终能否得到序列完整的 cDNA 文库. 传统的 cDNA 文库构建利用 mRNA 天然带有 poly(A)尾结构,使用 Oligo(dT)<sub>12-18</sub> 做逆转录得到 cDNA 第一链,再合成 cDNA 第二链并在双链 cDNA 两端加上适当的接头,然后连接到适当的载体中获得文库,但是,使用这种方法构建的 cDNA 文库会使得到的 cDNA 会比原始 RNA 序列少 5~30 个核苷酸<sup>[24]</sup>. 此外,在传统 sncRNA cDNA 文库构建中,T4 RNA 连接酶不能作用于 5'-端为非单磷酸结构的 sncRNA,显然影响 cDNA 文库的覆盖率.

本研究一方面保留了在 sncRNA 5'-端连接 RNA 接头的操作,这是因为 5'-接头只需 5'羟基(-OH)修饰,便能实现 5'-接头的定向连接,而且多数公司都能提供这种接头的合成;另一方面,由于多数 sncRNA 不带 poly(A)尾结构<sup>[15,16]</sup>,我们使用 *E. coli* Poly(A) Polymerase 对 sncRNA 3'-端加上 poly(A)尾,使 sncRNA 产生了类似 mRNA 的 poly(A)尾结构,就可将传统 mRNA cDNA 文库构建方法引入到 sncRNA cDNA 构建中. 另外,借鉴 3'-RACE(rapid amplification of cDNA ends)技术,使用带有两个简并核苷酸 VN 的接头引物 oligo(dT)<sub>16</sub>-VN-adapter-2 (V = A/G/C; N = A/G/

C/T)做逆转录获取 cDNA 第一链,能将引物定位在 poly(A)尾的起始点,消除了合成 cDNA 第一链时 oligo(dT)与 poly(A)尾配对的随机性。最后, Tobacco Acid Pyrophosphatase(TAP) 和 RNA 5' Polyphosphatase 的使用,拓展了建库起始 sncRNA 的类型,增加了 cDNA 文库覆盖率,如测序得到带帽结构的 RNA U2 的完整序列。将以上技术组合起来,改进了 sncRNA cDNA 文库构建方法,使 sncRNA cDNA 文库的完整性和有效性得到提高。

#### 参考文献:

- [1] Lander E S, Linton L M, Birren B, *et al.* Initial sequencing and analysis of the human genome [J]. *Nature*, 2001, 409: 860.
- [2] Sarah D, Carrie A D, Angelika M, *et al.* Landscape of transcription in human cells [J]. *Nature*, 2012, 489: 101.
- [3] Maria R, Angela C, Antonio R, *et al.* Small non-coding RNA deregulation in endometrial carcinogenesis [J]. *Oncotarget*, 2015, 6(7): 4677.
- [4] Young K K, Inha H, Narry K. Modifications of Small RNAs and Their Associated Proteins [J]. *Cell*, 2010, 143: 703.
- [5] Kawasaki H, Taira K. *MicroRNA*-196 inhibits *HOXB8* expression in myeloid differentiation of HL60 cells [J]. *Nucleic Acids*, 2004, 48: 211.
- [6] Robin C F, Kyle K F, Christopher B B, *et al.* Most mammalian mRNAs are conserved targets of microRNAs [J]. *Genome Research*, 2009, 19: 92.
- [7] Anthony K L, Phillip A S. MicroRNA Functions in Stress Responses [J]. *Molecular Cell*, 2010, 40: 205.
- [8] Alexei A, Dimos G, Sébastien P, *et al.* A novel class of small RNAs bind to MILI protein in mouse testes [J]. *Nature*, 2006, 442: 203.
- [9] Toshiaki W, Haifan L. Posttranscriptional Regulation of Gene Expression by Piwi Proteins and piRNAs [J]. *Molecular Cell*, 2014, 56:18.
- [10] Lasse P, Gunter M. Argonaute Proteins: Mediators of RNA Silencing [J]. *Molecular Cell*, 2007, 26: 611.
- [11] Panagiotis P, Juan V. The Spliceosome: The Ultimate RNA Chaperone and Sculptor [J]. *Trends Biochem Sci*, 2015, 41(1): 33.
- [12] Paul S, Scott P, Ruth S. RNA and DNA sequence recognition and structure-function of aminoacyl tRNA synthetases [J]. *Trends Biochem Sci*, 1982, 7(6): 209.
- [13] Monica R, John J R, Dominique L O. RNAi and small interfering RNAs in human disease therapeutic applications [J]. *Cell*, 2010, 28: 570.
- [14] Liao J P, Yu L, Mei Y P, *et al.* Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer [J]. *Molecular Cancer*, 2010, 9: 198.
- [15] Thomas D, Rory J, Giovanni B, *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression [J]. *Genome Research*, 2012, 22: 1775.
- [16] Nicole L G, Jeffrey W, Carol J W. The highways and byways of mRNA decay [J]. *Nat Rev Mol Cell Biol*. 2007, 8: 113.
- [17] Zhong S L, Joung J G, Zheng Y, *et al.* High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation [J]. *Cold Spring Harb Protoc*, 2011, 940.
- [18] Chen Y R, Zheng Y, Liu B, *et al.* A cost-effective method for Illumina small RNA-Seq library preparation using T4 RNA ligase1 adenylated adapters [J]. *Plant Methods*, 2012, 8: 41.
- [19] Wayne A S, Rebecca M T, Michael P T. The Box C/D motif directs snoRNA 5'-cap hypermethylation [J]. *Nucleic Acids Research*, 2000, 28(22): 4467.
- [20] Titia S, Florian A. S, Karen L, *et al.* Secondary siRNAs Result from Unprimed RNA Synthesis and Form a Distinct Class [J]. *Science*, 2007, 315: 244
- [21] Jackie R, Joost C Z. RNA-polymerase-I-directed rDNA transcription, life and works [J]. *Trends Biochem Sci*, 2005, 30(2): 87.
- [22] Louise C, John C. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome [J]. *Cell*, 1976, 9: 91.
- [23] Jesse M E, Amy P J, Patrick M, *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome [J]. *Science*, 2013, 341: 1237973.
- [24] Joseph S, David W R. *Molecular Cloning: A Laboratory Manual* [M]. 3rd ed. New York: Cold Spring Harbor Laboratory Press, 2001: 12. 1.