

Analysis of Circular RNA in *Caenorhabditis elegans*

Junwu Liu, Lingling Chen*

College of Informatics, Huazhong Agricultural University, Wuhan Hubei
Email: llchen@mail.hzau.edu.cn

Received: Jun. 11th, 2015; accepted: Jun. 25th, 2015; published: Jun. 30th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Exonic circular RNA (ecircRNA) is a class of special RNA molecule which is processed from pre-mRNA through back-to-splice. Recently, a lot of researches identified thousands of circular RNA in eukaryotes and found that circRNA could function as gene regulator. In our research, we predicted 1112 canonical circular RNAs in RNase R treated RNA-seq from different development stages of *Caenorhabditis elegans*. We discovered that these circular RNAs had developmental specificity, and had no obvious correlation with linear RNA. Through analyzing the sequence information of flanking introns, we inferred that circular RNAs in *C. elegans* were produced by intron-pairing-driven circularization model.

Keywords

Exonic Circular RNA, *C. elegans*, Developmental Specificity, Intron-Pairing-Driven Circularization

线虫环状RNA分析

刘骏武, 陈玲玲*

华中农业大学信息学院, 湖北 武汉
Email: llchen@mail.hzau.edu.cn

收稿日期: 2015年6月11日; 录用日期: 2015年6月25日; 发布日期: 2015年6月30日

*通讯作者。

摘要

外显子环状RNA (ecircRNA)是前体信使RNA中外显子经过反向剪接首尾相连而成环的一类特殊的RNA分子。近期多项研究发现,这类环状RNA广泛存在于真核生物细胞中,并具有调控基因表达的重要功能。本文根据线虫不同生长发育阶段的RNA测序数据,共预测出1112个典型的外显子环状RNA。对这些环状RNA进行分析,发现它们具有生长发育阶段特异性,并同线性的mRNA分子不具有直接的相关性。通过分析环状RNA侧翼内含子序列信息,证明在线虫中环状RNA的形成支持内含子配对驱动环化模型。

关键词

外显子环状RNA, 线虫, 生长发育特异性, 内含子配对驱动环化

1. 引言

外显子环状 RNA 指以 pre-mRNA 的外显子为主体,通过反向剪接的方式而头尾相连的一类特殊 RNA 分子(图 1)。外显子环状 RNA 最先发现可由人类的 *Ets-1*[1] [2]和小鼠的 *SRY*[3]基因产生,但是由于这一类环状 RNA 分子的表达量较低并且对它们的生物学功能缺乏足够的认识,因此被认为是由于剪切错误而形成的产物[4]。随着高通量测序技术和生物信息学的发展,在多种人类细胞组织[5]-[8]、小鼠[8]、果蝇[9]、线虫[8]等真核生物中以及古细菌[10]中预测出大量这类不带 poly(A)尾的环状 RNA 分子。近年研究表明,一些环状 RNA 可以作为 miRNA 分子海绵而大量结合 miRNA 分子[8] [11],从而阻碍 miRNA 对靶基因的调控作用,为一类新型的竞争性内源 RNA (competing endogenous RNA, ceRNA)。

环状 RNA 的 3'端和 5'端共价闭合,缺乏 poly(A)尾巴,一般认为不具有编码蛋白质功能。在细胞内比较稳定,并对核酸外切酶不敏感,可结合 miRNA 分子也可被 miRNA 切割降解[12]。外显子环状 RNA 通常定位于细胞质中,并认为环状 RNA 在形成过程中同时会将内含子剪切掉。而最近的研究表明,一些环状 RNA 的内含子会保留下来,并在细胞核中行使增强转录起始的功能[13]。

对于环状 RNA 的产生机制,Jeck 等[5]提出了环状 RNA 形成的两种模型:套索驱动环化(lariat-driven circularization)和内含子配对驱动环化(intron-pairing-driven circularization)。这两种模型的区别在于:套索驱动环化先由环状 RNA 两侧外显子的剪接供体 SD(splice donor)和剪接受体 SA(splice acceptor)共价结合形成套索结构,然后剪接体切除剩余内含子形成 circRNA;而内含子配对驱动环化则先由两侧内含子互补配对结合,然后剪接体切除内含子形成环状 RNA。近期的相关研究中[14]-[16],都支持了内含子配对驱动环化这一模型。

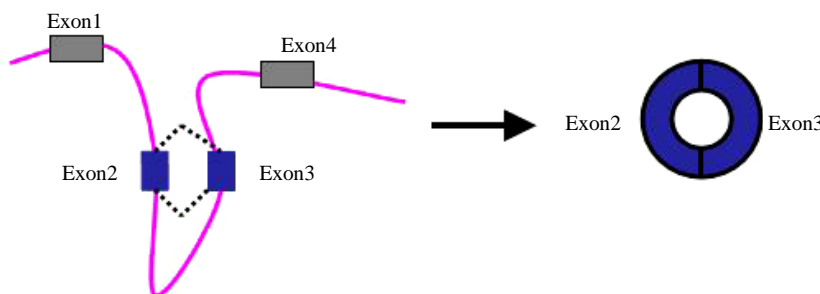


Figure 1. Formation of circular RNA

图 1. 环状 RNA 的形成

本文以模式生物线虫为例，通过经核酸外切酶 RNase R 富集后的 RNA-seq 数据，重新预测了线虫中存在的环状 RNA，并具体分析了环状 RNA 的一些基本特征，环状 RNA 同线性转录本的关系，环状 RNA 侧翼内含子序列间反向互补配对情况，环状 RNA 作为 miRNA 分子海绵的可能性，并对产生环状 RNA 的基因进行 GO 功能注释。以期为其他物种环状 RNA 及非编码 RNA 的研究提供一些参考信息。

2. 材料和方法

2.1. 数据来源

线虫 RNA-seq 数据来自 NCBI GEO datasets 数据库中的 GSE63823[16]，为一组线虫不同生长发育阶段：胚胎期(embryo)、幼虫期(L1~L4)、成虫期(adult)的 total RNA(rRNA-deplete)及对应时期经核酸外切酶 RNase R 处理后的 RNA-seq 数据，测序数据信息统计如表 1 所示。

线虫参考基因组(Ce6)及基因组注释信息从数据库 UCSC[17]中下载得到。

线虫 miRNA 成熟体序列从 miRNA 收录数据库 miRBase (v21) [18]中下载得到。

2.2. 数据分析方法

2.2.1. 数据预处理

使用 FastQC (v0.10.1)(<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>)对测序数据进行质量检测，确保质量可以进行后续的分析；如果测序质量不好，可用 Trimmomatic (v0.30)[19]进行质量控制。

2.2.2. 环状 RNA 的预测

本文使用 find_circ 流程[8]对线虫环状 RNA 进行预测。find_circ 流程首先将不能和基因组比对上的读段两端各取 20 bp 作为锚点，再将锚点作为独立的读段往基因组上比对并寻找唯一匹配位点，如果两个锚点的比对位置在线性上方向呈反向，就延长锚点的读段，直至找到环状 RNA 的接合位置，若此时两侧的序列分别为 GT/AG 剪接信号，则判断为环状 RNA。相关程序可在 circBase[20]数据库中下载得到。预测步骤如下：

Table 1. The RNA-seq datasets of *C. elegans*

表 1. 线虫 RNA-seq 数据集

| Datasets | Stage | Treatment | Coverage |
|------------|--------|-----------|----------|
| GSM1558150 | embryo | None | ~29.9 x |
| GSM1558151 | embryo | RNase-R | ~69.3 x |
| GSM1558152 | L1 | None | ~51.8 x |
| GSM1558153 | L1 | RNase-R | ~64.4 x |
| GSM1558154 | L2 | None | ~38.5 x |
| GSM1558155 | L2 | RNase-R | ~61.5 x |
| GSM1558156 | L3 | None | ~44.7 x |
| GSM1558157 | L3 | RNase-R | ~63.3 x |
| GSM1558158 | L4 | None | ~46.4 x |
| GSM1558159 | L4 | RNase-R | ~54.8x |
| GSM1558160 | adult | None | ~41.4 x |
| GSM1558161 | adult | RNase-R | ~65.9 x |

- 1) 基因组建索引: bowtie2-build genome.fa genome.fa(bowtie2v2.2.1)。
- 2) 安装 python 包: numpy (<http://www.numpy.org/>)和 pysam (<https://pypi.python.org/pypi/pysam>)。
- 3) 建立一个文件夹 genome, 里面存放基因组中各个染色体的 fasta 文件, 文件命名需同基因组 fasta 文件里的 id 相同。
- 4) 将处理后的读段比对到参考基因组上并将结果排序和转换成 BAM 格式: bowtie2 -p 20 -very-sensitive --score-min=C,-15,0 -q -x genome.fa -1sample_1.fastq -2 sample_2.fastq 2>log/sample.bowtie2.log| samtools view -hbuS - |samtools sort - sample。
- 5) 取出未匹配上的读段: samtools view -hf 4 sample.bam|samtools view -Sb -> BAM/unmapped_sample.bam。
- 6) 将未匹配的读段两端各取 20 bp 作为锚点序列: python unmapped2anchors.py BAM/unmapped_sample.bam|gzip>gzip/sample_anchors.qfa.gz。
- 7) 将锚点同基因组比对, 并得到可能的环状 RNA 位点: bowtie2 -p 23 --reorder --score-min=C,-15,0 -q -x genome.fa -U gzip/sample_anchors.qfa.gz |python find_circ.py -G genome -p sample -s log/find_circ.log >bed/sample.sites.bed 2>bed/sample.sites.reads。
- 8) 筛选候选环状 RNA: grep circ bed/sample.sites.bed | grep -v chrM | python sum.py -2,3 |python scorethresh.py -16 1|python scorethresh.py -15 2|python scorethresh.py -14 2 |python scorethresh.py 7 2 |python scorethresh.py 8,9 35 |python scorethresh.py -17 100000 > bed/sample.circ_candidates.bed。

2.2.3. 表达量统计

Cufflinks (v2.2.0) [21]是一款常用的转录本组装工具, 它利用 Tophat (v2.0.11) [22]比对的结果来组装转录本, 估计这些转录本的丰度并且可以检测样本间的表达差异及可变剪接。

利用 FPKM[23] (Reads per kilobase of exon model per Million mapped reads)归一化各个转录本的表达量。

FPKM 计算公式:

$$\text{FPKM} = \frac{\text{total exon fragment}}{\text{mapped reads (Millions)} \times \text{exon length (kb)}}$$

流程如下:

- 1) 基因组建索引: bowtie2-build genome.fa genome.fa;
- 2) Tophat 比对: tophat -a 8 -m 0 -I 300000 -p 20 --library-type fr-unstranded --no-novel-indels --segment-mismatches 2 -G genome.gtf -o tophat/sample genome.fa sample.fastq
- 3) Cufflinks 组装转录本并估计表达量: cufflinks -p 20 -G genome.gtf --library-type fr-unstranded -o cufflinks/sample tophat/sample/accepted_hits.bam

我们选取 Y45F10D.4 为管家基因[24], 用下面的归一化公式计算环状 RNA 的表达量:

$$\text{Exp} = \frac{\text{junction spanning reads/million raw reads}}{\text{housekeeping gene FPKM}}$$

2.2.4. miRNA 结合位点统计

我们使用滑动窗口的方法确定线虫各个 miRNA 在环状 RNA 上的结合位点数, 窗口大小为 7 bp, 窗口的滑动步长为 1 bp, 先用 perl 脚本提取线虫成熟体 miRNA 的种子序列(第 2~8 nt), 然后每滑动一次窗口, 判断该窗口是否同种子序列反向互补, 若反向互补, 则判定为该 miRNA 的一个结合位点。

2.2.5. 内含子反向互补配对

根据内含子互补配对驱动环化模型，我们使用 blast 将环状 RNA 的上下游内含子序列进行比对，为了尽可能使 blast 2.2.29+ 结果倾向于更长的匹配，我们设定参数为：-ungapped -word_size 6 -penalty -1，编写 perl 脚本统计内含子内部和内含子间的反向互补配对数目。

3. 结果与分析

3.1. 线虫环状 RNA 的基本特征

我们共鉴定出 1531 个环状 RNA，舍去 79 个位于线虫基因组 rRNA 簇 chrI: 15,060,128~15,071,020 上的环状 RNA，共 1452 个作为候选的 circRNA。根据 circRNA 的供体和受体所处的位置(图 2)，统计得出：

- 1) canonical: circRNA 的起始和终止点都在相同基因的外显子边界上，共 1112 个；
- 2) alt_donor: circRNA 的起始位点在外显子边界上而终止位点不在，共 68 个；
- 3) alt_acceptor: circRNA 的终止位点在外显子边界上而起始位点不在，共 65 个；
- 4) antisense: circRNA 与基因区重叠且处于反义链上，共 20 个；
- 5) intergenic: circRNA 与基因间区重叠，共 162 个；
- 6) other: 其他情况，共 25 个。

在这 1452 个环状 RNA 中，仅 119 个在线虫的各个时期中均有表达。在某一特定时期表达的环状 RNA 数目相对于其他时期而言，具有相同环状 RNA 的比例都不高(表 2)，说明线虫环状 RNA 在不同生长发育阶段会表达不同的环状 RNA，具有生长发育阶段特异性。

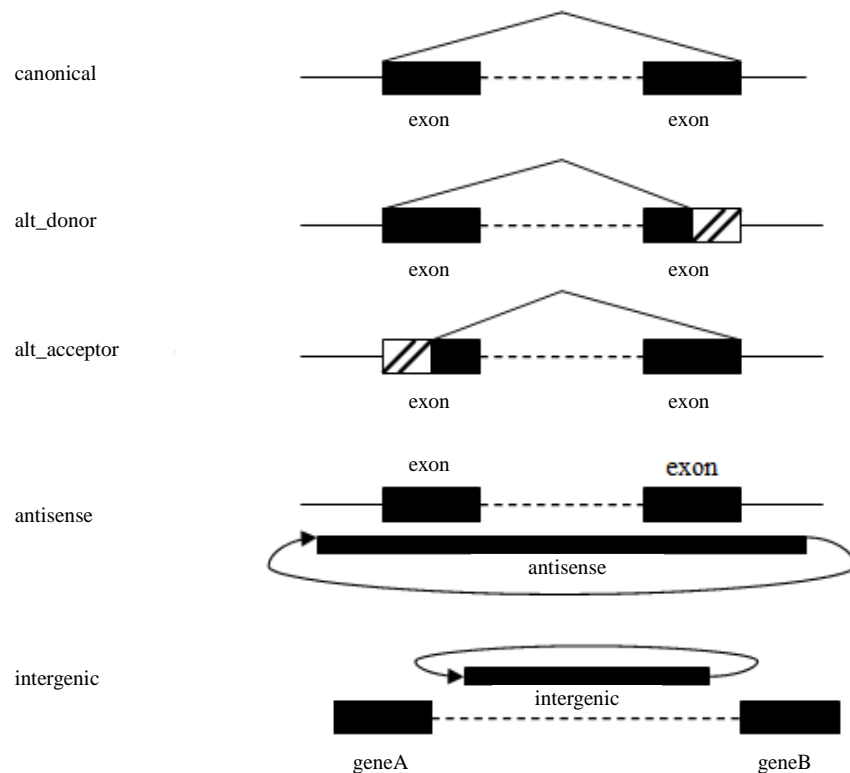


Figure 2. Type of circular RNA
图 2. 环状 RNA 的类型

Table 2. The number of the same circular RNA among different stages
表 2. 各个生长时期相同的环状 RNA 数目

| Stage same | embryo | L1 | L2 | L3 | L4 | adult |
|---------------|--------|-----|-----|-----|-----|-------|
| embryo | 592 | | | | | |
| L1 | 286 | 716 | | | | |
| L2 | 305 | 397 | 683 | | | |
| L3 | 323 | 381 | 422 | 693 | | |
| L4 | 159 | 168 | 179 | 192 | 224 | |
| adult | 241 | 243 | 259 | 272 | 153 | 367 |

此外,我们将预测出的 1112 个 canonical 环状 RNA 映射回基因上,由于一个基因的选择性剪切会形成多个不同组成的转录本,我们无法判断环状 RNA 真实来自于哪一个转录本,所以我们假设只要含有该环状 RNA 边界外显子的转录本都会形成,那么我们一共可以对应到 2056 个转录本。我们对这些环状 RNA 使用的外显子数目和长度分布进行统计,可以看到环状 RNA 由中间外显子组成的比例最大(90.08%),并且值得注意的是由第一个外显子的数目组成的环状 RNA 有 203 个(图 3(a)),这个数目和比例远远超过了人类环状 RNA 的情况,而环状 RNA 形成的两个模型中也并没有描述由第一个外显子形成环状 RNA 的情况;由两个外显子组成的环状 RNA 数目最多,其次为由三个外显子和一个外显子组成的情况(图 3(c)),包含外显子越多的环状 RNA 数目越少,最多的环状 RNA 可由 12 个外显子构成(ce1_circ_F08B4.2a 和 ce1_circ_F08B4.2b);对比组成环状 RNA 的外显子长度可以发现,在只由一个外显子组成的环状 RNA 中,外显子的中位数长度达到了 333 bp,要明显高于由多个外显子组成的环状 RNA 中的外显子长度($p < 0.0001$) (图 3(c)),因此我们推测,环状 RNA 需要达到一定长度才会形成。另外,我们统计了环状 RNA 的长度分布,最短环状 RNA 长度为 116 bp,最长为 4131 bp,中位数长度为 395 bp (图 3(d))。

我们将环状 RNA 对应的转录本和环状 RNA 之间的表达量进行相关性分析,发现在 97 个在所有时期都表达的 canonical 环状 RNA 中,仅有 11 个环状 RNA 同其线性转录本相关系数大于 0.8,1 个小于负 0.8。表明多数环状 RNA 同线性 RNA 表达量间并没有明显的相关性,环状 RNA 并不是简单地由线性 RNA 中的一部分形成,它还受到其他多种因素的调控,呈现出复杂的调控机制。

3.2. RNase R 对环状 RNA 丰度的影响

RNase R 是一种 3' → 5' 核酸外切酶,可以切割线性的 RNA 分子,因此认为它在 mRNA 降解过程中扮演着重要的角色[25]。在环状 RNA 的研究中,认为环状 RNA 是对 RNase R 不敏感的,因此 RNase R 的处理会对环状 RNA 起富集作用,并且为判定 RNA 是否成环的一个重要条件。在本文的实验中,我们通过对前后经过 RNase R 处理的 RNA-seq 数据,虽然 RNase R 处理后的 RNA-seq 数据深度提高了,但经过归一化比较后发现大部分环状 RNA 在经过 RNase R 处理后的确是富集了,但是仍有一部分没有富集,还有一小部分环状 RNA 对 RNase R 敏感(图 4)。

在对 RNase R 敏感的环状 RNA 中,大部分是 intergenic 类型,图 4 所展示的 chrX: 925,367~926,302 是唯一表现出敏感的 canonical 类型的环状 RNA。而目前尚没有关于 intergenic 类型环状 RNA 的研究,故不能确定这一类环状 RNA 是否真的存在并对 RNase R 敏感还是软件预测的假阳性。

3.3. 环状 RNA 的多样性和可选择性

环状 RNA 可以由一个外显子或多个外显子构成,一个基因中也会有形成多个环状 RNA 的情况:环

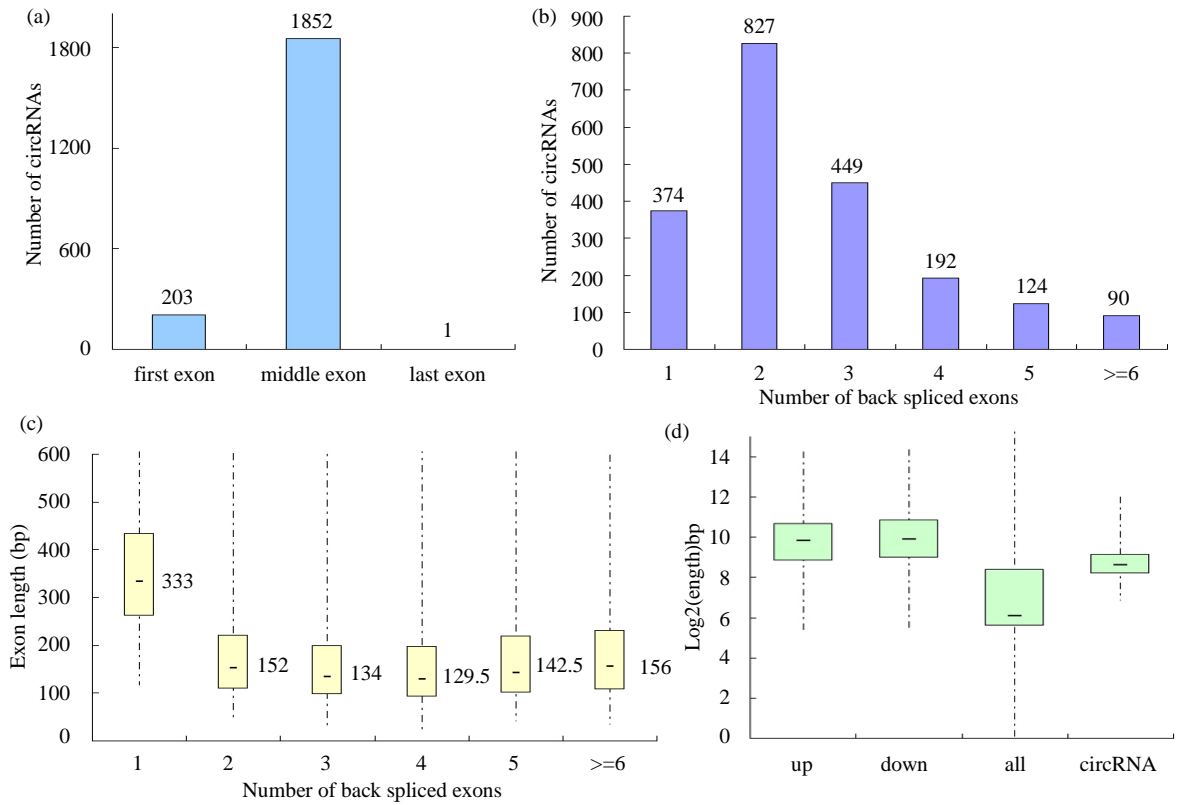


Figure 3. Basic feature analysis of circular RNAs in *C.elegans*
图 3. 线虫环状 RNA 的基本特征分析

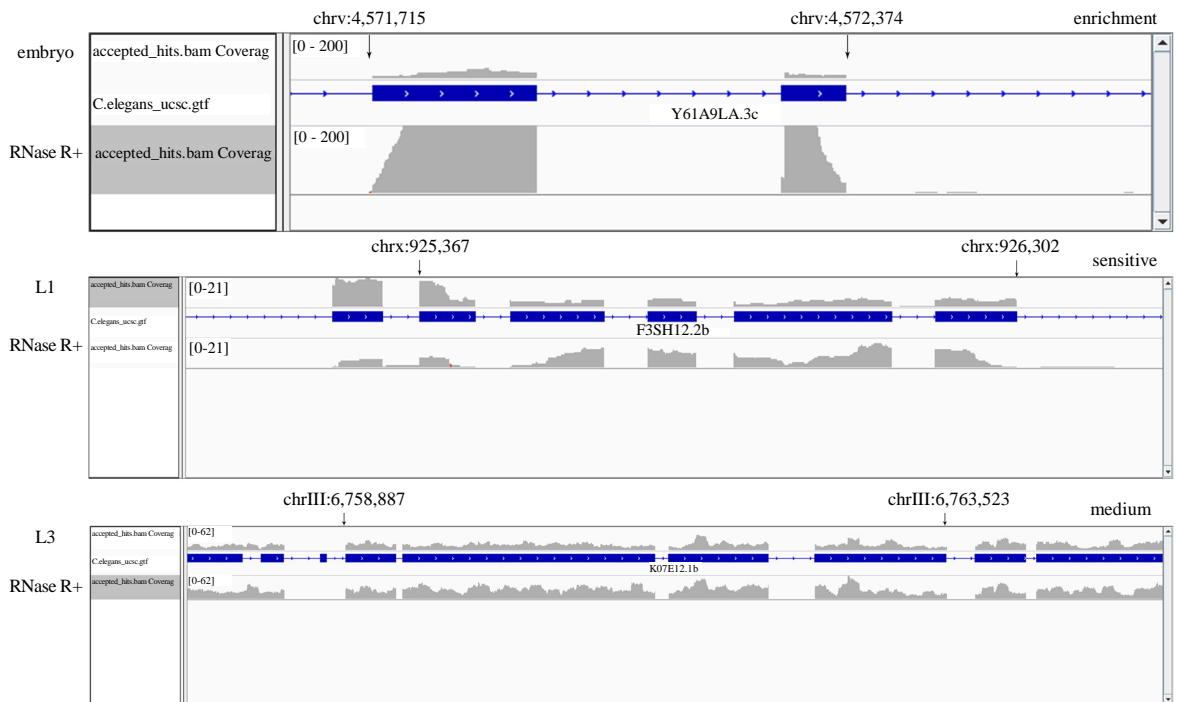


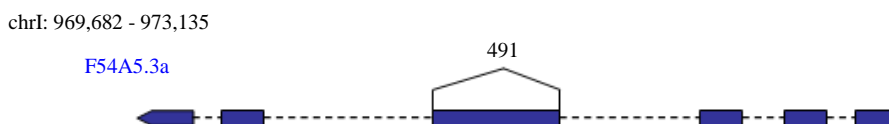
Figure 4. Read counts variation before and after RNase R treatment: enrichment, sensitive and medium
图 4. RNase R 处理前后环状 RNA 读数变化：共富集、敏感、中等三种情况

状 RNA 可能由基因的不同外显子组成，也可能供体或受体为相同的外显子，也可能多种情况在一个基因中同时发生，不同的可变剪接体间形成的环状 RNA 还会竞争外显子等情况(图 5)。环状 RNA 对外显子具有选择性，称为可变化[14]。

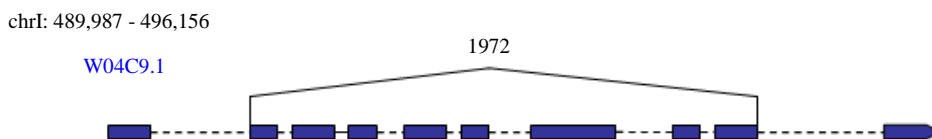
3.4. 线虫环状 RNA 侧翼内含子特征

线虫环状 RNA 上游和下游内含子的中位数长度分别达到了 907 bp 和 962 bp，它们之间没有显著差异($p = 0.157$)，但同线虫所有内含子的中位数长度(69 bp)有极其显著差异($p < 0.0001$) (图 3(d))，说明具有较长的侧翼内含子序列是环状 RNA 的一个基本特征。

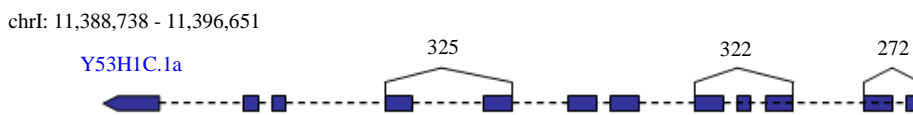
单个外显子环化



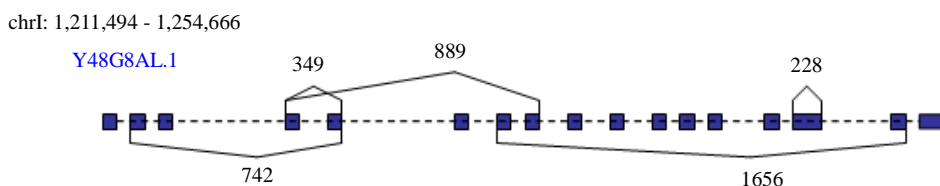
多个外显子环化



基因内可发生多次环化



复杂和可选择性的环化方式



可变剪接体间环化的竞争

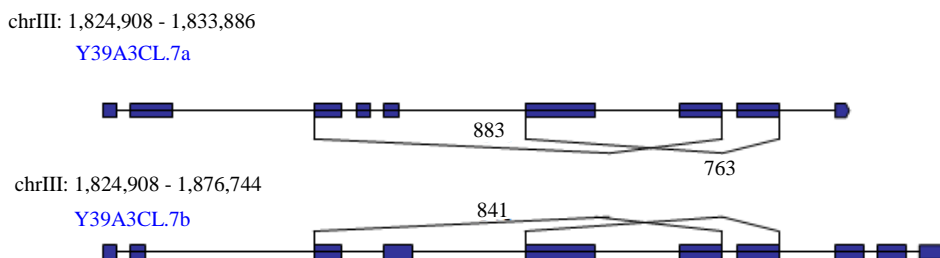


Figure 5. Examples that illustrate the diversity of circularizing exons

图 5. 线虫环状 RNA 多样的环化方式

Alu 元件是人类基因组上大量存在的短散在重复序列，也是人类环状 RNA 形成的主要原因，同时环状 RNA 两侧的内含子序列内(IRwithin)和序列间(IRacross)的反向互补配对的竞争(图 6(a))也是造成环状 RNA 形成的重要原因[14]。由于在线虫基因组中 83%都是非重复序列，因此线虫环状 RNA 形成的主要原因不会是因为基因组上的重复序列。我们根据环状 RNA 的位点信息，提取出环状 RNA 侧翼内含子序列，并用 blast 来统计 IRwithin 和 IRacross 的情况。我们发现 96.33%的环状 RNA 侧翼内含子都存在序列反向互补情况(图 6(b))，只有约 0.3%的内含子间不含有 IRacross，在 1852 个序列对中，IRacross - IRwithin ≥ 1 的有 1374 对，占 74.19%。说明具有反向互补配对的侧翼内含子序列也是线虫环状 RNA 的一个重要的基本特征。

同时我们也统计了 203 个由第一个外显子构成的环状 RNA 的上游基因间区序列和其下游第一个内含子之间的序列比对情况，发现具有 IRacross 的有 180 对，占 88.67%，IRacross - IRwithin ≥ 1 的有 107 对，占 52.71%(图 6(b))。真核基因转录产物一般是单顺反子，但线虫中约 25%的基因会产生多顺反子，因此我们推测这些顺反子之间的基因间区可能会和 intron 区的序列反向互补配对，从而造成第一个外显子也参与成环的现象。

3.5. miRNA 在环状 RNA 上的结合位点

在 canonical 和 antisense 类型的环状 RNA 中，每个环状 RNA 都有一个或多个 miRNA 的靶位点，最多的可以有 95 个不同 miRNA 的靶位点，但可结合相同 miRNA 的数目却大多只有 1~3 个。其中环状 RNA 位点(chrIII: 6758887-6763523)可由 *K07E12.1a.1*、*K07E12.1a.2*、*K07E12.1b* 三个不同转录本中的三个外显子形成，该环状 RNA 对 *cel-miR-77-3p* 有 16 个结合位点，跟 *SRY* 基因对 miR-138 结合的数目相同[11]，故该 circRNA 可能具有潜在的 miRNA 分子海绵作用，而其他的环状 RNA 由于可结合相同 miRNA 的数目太少，因此认为它们不具有分子海绵的功能。

我们的结果中可以看出，在线虫中环状 RNA 可作为分子海绵的数量和能力十分有限，而在目前已报导的 circRNA 中，也只有 *CDR1as* 和 *SRY* 两个环状 RNA 被证明具有分子海绵作用。

3.6. GO 功能富集

我们将可以产生环状 RNA 的转录本进行 GO 功能富集分析[26] (图 7)，GO 富集结果表明多数基因在细胞成分上为细胞和细胞器成分，具有结合蛋白，催化，运输等分子功能，参与到生物调节，细胞过程，生长过程，免疫，代谢等多种生物过程中。

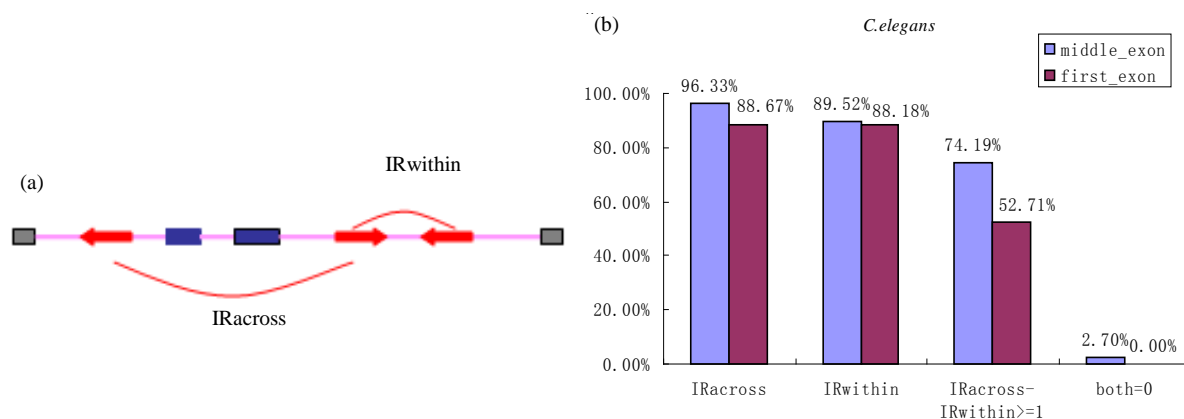


Figure 6. Reverse complementary matches of circular RNA flanking intron in *C.elegans*

图 6. 线虫环状 RNA 侧翼内含子序列反向互补情况

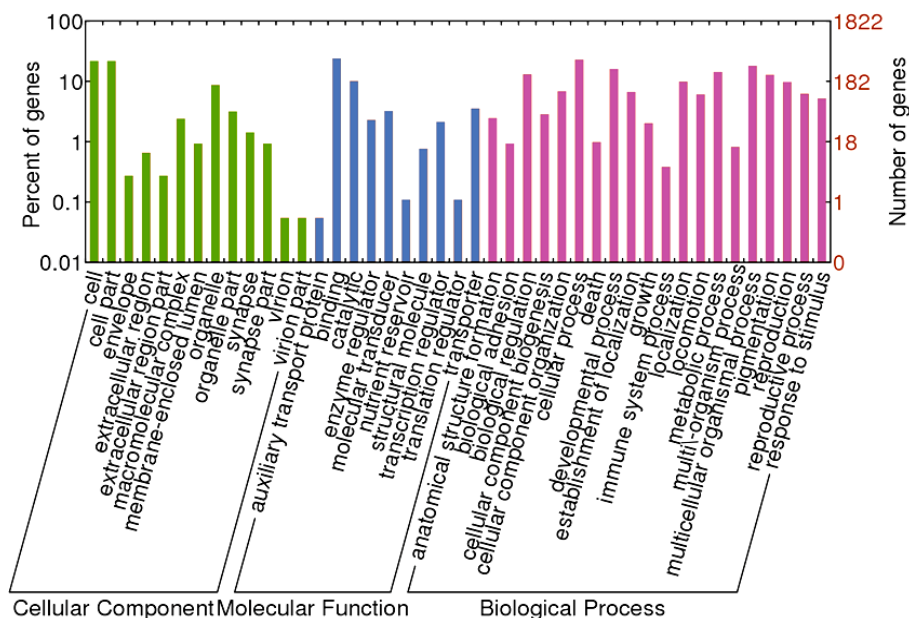


Figure 7. GO annotation for genes which can generate circular RNA
 图 7. 可产生环状 RNA 的基因 GO 功能注释

4. 讨论

本文根据核酸外切酶 RNase R 处理后的 RNA-seq 测序数据，重新预测线虫中存在的环状 RNA，数目由原先的 725 个提升到 1452 个。在得到的环状 RNA 中，有一些经过 RNase R 处理后读段降低或消失的其他类型环状 RNA。通过对环状 RNA 的构成和基本特征进行分析，发现线虫中环状 RNA 主要由中间两个外显子构成，但是由第一个外显子形成环状 RNA 的比例较高，我们认为原因是由于线虫可以产生多顺反子。通过分析环状 RNA 侧翼内含子序列，发现线虫环状 RNA 内含子长度较长且上下游区间多含有反向互补配对序列，同样支持了内含子配对驱动环化模型。

在环状 RNA 作为 miRNA 分子海绵的可能性上，线虫中虽然每个环状 RNA 上都有 miRNA 的结合位点，但我们只发现一个环状 RNA 可以结合 16 个相同的 miRNA 分子，可能为潜在的分子海绵，而其他的环状 RNA 都没有 10 个以上的 miRNA 结合位点。此外，我们还检测了人类 *CDR1as* 序列上其他 miRNA 分子的结合位点，发现有 119 个 miR-5011-5p 位点、116 个 miR-190a-3p 位点、84 个 miR-6873-3p 位点，这些都比 miR-7(74 个)的数量要多，但是却没有任何报导 *CDR1as* 可以作为这些 miRNA 的分子海绵。说明只有少数的环状 RNA 可以充当特定 miRNA 的分子海绵。

目前关于环状 RNA 功能的研究较少，有研究表明在果蝇和小鼠的脑组织中环状 RNA 表达的丰度会高于其他组织，环状 RNA 可能和神经发育相关[9] [27]，并且 *CDR1as* 的表达也会阻碍中脑的发育。尽管如此，环状 RNA 究竟为什么而产生，还可以行使怎样的功能尚值得我们进一步研究。而准确的建库方式以及专门寻找环状 RNA 的生物信息学工具[28]的开发会使我们更好地研究这类新型的 RNA 分子。

参考文献 (References)

- [1] Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) Scrambled exons. *Cell*, **64**, 607-613.
- [2] Cocquerelle, C., Daubersies, P., Majerus, M.-A., Kerckaert, J.-P. and Bailleul, B. (1992) Splicing with inverted order of exons occurs proximal to large introns. *The EMBO Journal*, **11**, 1095-1098.
- [3] Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. (1993)

- Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, **73**, 1019-1030.
- [4] Cocquerelle, C., Mascrez, B., Hetuin, D. and Bailleul, B. (1993) Mis-splicing yields circular RNA molecules. *The FASEB Journal*, **7**, 155-160.
- [5] Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F. and Sharpless, N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141-157.
- [6] Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L. and Brown, P.O. (2013) Cell-type specific features of circular RNA expression. *PLoS Genet*, **9**, e1003777.
- [7] Guo, J.U., Agarwal, V., Guo, H. and Bartel, D.P. (2014) Expanded identification and characterization of mammalian circular RNAs. *Genome Biology*, **15**, 304-313.
- [8] Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F. and Rajewsky, N. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333-338.
- [9] Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R. and Lai, E.C. (2014) Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Reports*, **9**, 1966-1980.
- [10] Danan, M., Schwartz, S., Edelheit, S. and Sorek, R. (2012) Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Reserach*, **40**, 3131-3142.
- [11] Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384-388.
- [12] Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. and Kjems, J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO Journal*, **30**, 4414-4422.
- [13] Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., Zhu, P., Chang, Z., Wu, Q., Zhao, Y., Jia, Y., Xu, P., Liu, H. and Shan, G. (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural Molecular Biology*, **22**, 256-264.
- [14] Zhang, X.O. and Wang, H.B., Zhang, Y., Lu, X., Chen, L.L., Yang, L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134-147.
- [15] Liang, D. and Wilusz, J.E. (2014) Short intronic repeat sequences facilitate circular RNA production. *Genes and Development*, **28**, 2233-2247.
- [16] Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C. and Rajewsky, N. (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Reports*, **10**, 170-177.
- [17] Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Research*, **12**, 996-1006.
- [18] Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, **42**, D68-D73.
- [19] Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
- [20] Glažar, P., Papavasileiou, P. and Rajewsky, N. (2014) circBase: A database for circular RNAs. *RNA*, **20**, 1666-1670.
- [21] Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562-578.
- [22] Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, **14**, R36.
- [23] Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**, 621-628.
- [24] Zhang, Y., Chen, D., Smith, M.A., Zhang, B. and Pan, X. (2012) Selection of reliable reference genes in *Caenorhabditis elegans* for analysis of nanotoxicity. *PLoS ONE*, **7**, e31849.
- [25] Cheng, Z.-F. and Deutscher, M.P. (2005) An important role for RNase R in mRNA decay. *Molecular Cell*, **17**, 313-318.
- [26] Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R. and Bolund, L. (2006) WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, **34**, W293-W297.
- [27] You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C.,

- Wang, X., Hou, J., Liu, H., Sun, W., Sambandan, S., Chen, T., Schuman, E.M. and Chen, W. (2015) Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neuroscience*, 18, 603-610.
- [28] Gao, Y., Wang, J. and Zhao, F. (2015) CIRI: An efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology*, **16**, 4.