Revealing the diverse functional roles of noncoding RNA

Summary

- Many types of noncoding RNAs exist, such as ribosomal RNA, transfer RNA, long noncoding RNA, and a variety of microRNAs
- Noncoding RNAs can have important biological functions in gene regulation and disease development, yet the roles of the majority of noncoding RNAs remain largely unknown
- Whole-transcriptome sequencing provides a comprehensive analysis to identify and explore the function of noncoding RNAs

Introduction

A total RNA sample is composed of multiple types of transcripts, including coding and noncoding RNAs. Messenger RNA (mRNA), the coding portion of total RNA, accounts for only about 1–4% of the total RNA population, and the remainder is generally considered to be noncoding. Noncoding RNA (ncRNA) is defined as RNA that is not translated into protein, and there are many types of ncRNA with a variety of biological functions [1] (Figure 1). These types include ribosomal RNA (rRNA), transfer RNA (tRNA), long ncRNA (lncRNA; transcripts longer than 200 nucleotides not translated into protein), and many smaller ncRNAs such as microRNA (miRNA), small nuclear RNA (snRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), small nucleolar RNA (snoRNA), and small Cajal body–specific RNA (scaRNA). A DNA sequence encoding an ncRNA is often called an “RNA gene”. The relative amounts of the different types of ncRNAs vary greatly among species and cell types. Most prevalent is rRNA, which typically accounts for 80–95% of the total RNA population. The remainder of ncRNAs are present in much smaller amounts, and thus may require larger samples or enrichment procedures in order to acquire enough material to study. Currently, thousands of ncRNAs have been identified, and it is thought that there are many more yet to be discovered [1].

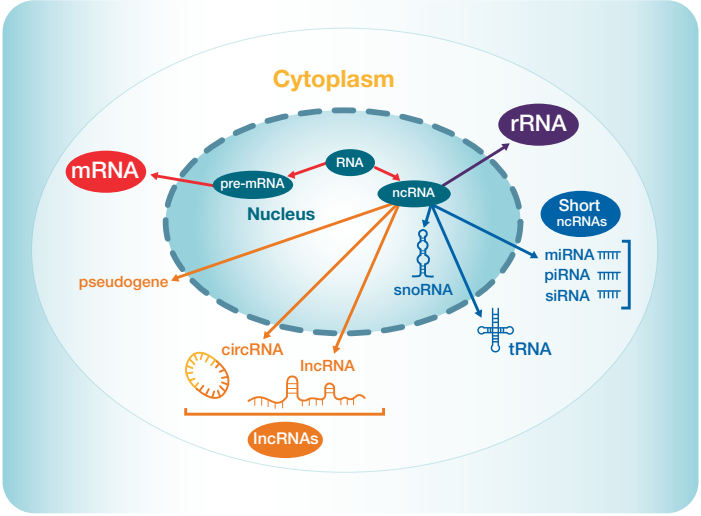

Figure 1. Types of noncoding RNA and their respective locations within a typical eukaryotic cell.

ThermoFisher SCIENTIFIC

ncRNAs can have crucial roles in various biological processes. The majority of ncRNAs appear to be involved in transcriptional or translational processes, and they are important in both healthy and disease states. Examples include inhibiting translation, maintaining the integrity of chromosomes, regulating protein activity, causing mRNA degradation, forming a variety of ribonucleoprotein complexes like splicesomes, or even functioning as transcription coactivators [2]. Autism, Alzheimer's disease, certain types of cancers, and Prader-Willi syndrome are some examples of diseases thought to be caused by the dysfunction of different types of ncRNAs. miRNAs, which regulate gene expression by binding to their target mRNAs, are implicated in the development of acute lymphoblastic leukemia [3,4]. Prader-Willi syndrome is a developmental disorder that is caused by deletions of the C/D box snoRNA SNORD116 [5].

Studying ncRNA function

Efficient methods are available to identify and study the function of ncRNAs. The detection of ncRNAs depends on a careful analysis of the transcriptome. NGS-based methods, such as whole-transcriptome sequencing with RNA-Seq, can be used to examine ncRNAs. RNA-Seq can characterize novel RNA species and analyze gene expression patterns rapidly and cost-effectively [6]. This technology can be used to locate where in the genome ncRNAs originate, where in the cell they may function, and also in which conditions or processes they may be involved.

Pseudogenes

Pseudogenes are a unique category of DNA sequences that may have biological importance. They are copies of a parent gene that contain premature stop codons or frameshift mutations. Typically, pseudogenes are predicted based on similarities to known functional genes, and they can produce ncRNAs on occasion. For example, the pseudogene Lethe produces a lncRNA that can inhibit the ability of RelA to bind to NF-κB gene promoters [7]. In addition, some pseudogenes are translated, and their expression tends to be restricted to certain tissues or cells, as opposed to being ubiquitous—underscoring the evidence for possible functional roles. A detailed analysis of the human proteome found that more than 200 peptides encoded by 140 pseudogenes were translated, yet the functions of the protein products remain largely unknown [8].

Transcription start sites

The determination of transcription start sites (TSSs) is important because there are many more protein products than genes, and this is the result of alternative TSSs that generate isoforms. Mapping of alternative TSSs is needed to understand isoform diversity, as well as gene regulation in general. The identification of intergenic TSSs involves a detailed analysis of the transcriptome, which can aid in the identification of novel ncRNAs and determination of their biological relevance. Alternative TSSs can be used under different cellular conditions—in the presence of disease or during different developmental stages—and they can exhibit tissue or cell specificity. RNA-Seq can provide a rapid genome-wide method to determine TSSs, enabling a comparison between different cell types, among a variety of conditions, and even during different disease states.

ncRNA databases and resources

Many high-quality ncRNA databases exist, underscoring their importance and the rapid increase in knowledge in this field (Table 1). Most of these resources provide open access, and they can include data from multiple species of ncRNA from many different organisms. Rapid advances in technology are enabling the identification and functional evaluation of thousands of new ncRNAs, and it is expected that more databases and resources will be needed in the near future to organize and store all of this information.

Table 1. Examples of ncRNA databases.

Database	Focus area	Access version
NONCODE	Integrated knowledge database dedicated to ncRNAs (excluding tRNAs and rRNAs)	v5.0 (Aug 2018)
RNAcentral	Central database for a range of ncRNA databases	Assorted
lncRNome	Biologically oriented knowledgebase for lncRNAs in humans	GENCODE v12 (Aug 2018)
Lnc2Cancer	Database of experimental support for associations between lncRNA and human cancer	v2.0 (Aug 2018)
Nervous System Disease NcRNAome Atlas	Database of experimental support for associations between ncRNA and nervous system diseases (NSD)	May 2017

Results

To determine the representation of different noncoding RNA species in a range of RNA library preparation methods, RNA libraries were prepared from 1,000 ng of human brain reference RNA according to manufacturer instructions, and then sequenced on an Illumina™ HiSeq™ 4000 system. The representation of the different RNA species was determined by analysis with QoRTs software [9] (Figure 2). The Invitrogen™ Collibri™ Stranded RNA Library Prep Kit for Illumina and NEBNext™ Ultra™ II RNA Library Prep Kit provided the best coverage of lncRNA. The most abundant types of lncRNA represented in the library produced using the Collibri kit were antisense RNA and long intergenic noncoding RNA (lincRNA). For short ncRNA, the library prepared with the Collibri kit included snRNA, snoRNA, and miRNA (Figure 2); in particular, snoRNA reads were better represented than with the other library preparation methods tested.

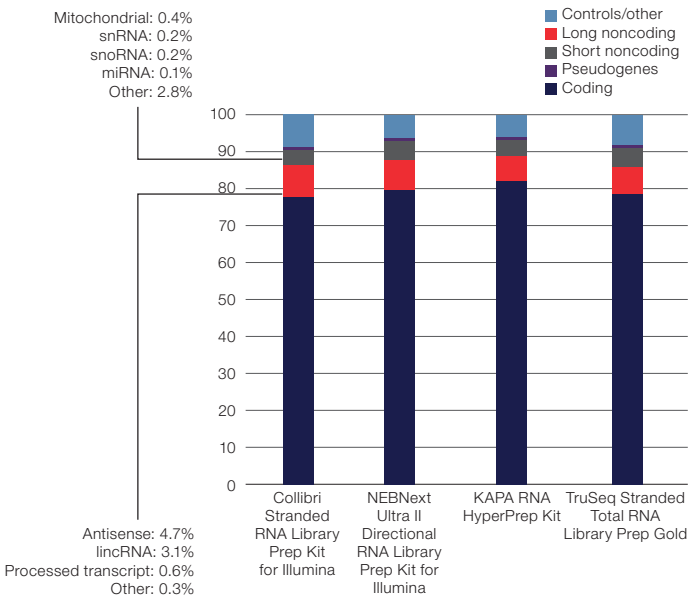

Figure 2. Representation of ncRNA in four different library preparation methods. RNA libraries were prepared from 1,000 ng of human brain RNA according to manufacturer instructions. The representation of the different RNA species was determined by analysis with QoRTs software.

Conclusions

- Thousands of ncRNAs have already been identified; some are biologically important, yet it is likely that the majority have yet to be characterized
- Characterization of ncRNAs relies upon a detailed analysis of the transcriptome, which can be performed most efficiently with an NGS-based approach
- RNA-Seq can provide a comprehensive analysis of the transcriptome, enabling the discovery of novel ncRNAs and their biological roles in both health and disease

References

1. Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? *Front Genet* 6:2.
2. Fu X-D (2014) Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev* 1:190-204.
3. Ultimo S, Martelli AM, Zauli G et al. (2017) Roles and clinical implications of microRNAs in acute lymphoblastic leukemia. *J Cell Physiol* 233:5642-5654.
4. Luan C, Yang Z, Chen B (2015) The functional role of microRNA in acute lymphoblastic leukemia: relevance for diagnosis, differential diagnosis, prognosis, and therapy. *Onco Targets Therapy* 8:2903-2914.
5. Sahoo T, del Gaudio D, German JR et al. (2008) Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 40:719-721.
6. Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA* 8(1).
7. Rapicavoli NA, Qu K, Zhang J et al. (2013) A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* 2:e00762.
8. Kim M-S, Pinto SM, Getnet D et al. (2014) A draft map of the human proteome. *Nature* 509:575-581.
9. Hartley SW, Mullikin JC (2015) QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* 16:224.

invitrogen

Ordering information

Product	Quantity	Cat. No.
RNA-Seq kits		
Collibri Stranded RNA Library Prep Kit for Illumina*	24 preps	A38994024
	96 preps	A38994096
Collibri Stranded RNA Library Prep Kit for Illumina with H/M/R rRNA Depletion Kit*	24 preps	A39003024
	96 preps	A39003096
ERCC RNA Spike-In Mix	1 kit	4456740
ERCC ExFold RNA Spike-In Mixes	1 kit	4456739
Library quantification		
Collibri Library Quantification Kit*	100 rxns	A38524100
	500 rxns	A38524500
Qubit 4 Fluorometer	1 instrument	Q33238
Qubit 1X dsDNA HS Assay Kit	100 assays	Q33230
Qubit 4 NGS Starter Kit	1 kit	Q33240
Library amplification		
Platinum SuperFi Library Amplification Master Mix	50 rxns	A38539050
	250 rxns	A38539250
Platinum SuperFi Library Amplification Master Mix with Primer Mix*	50 rxns	A38540050
	250 rxns	A38540250
Purification		
PureLink RNA Mini Kit	25 preps	12183020
	250 preps	12183025
MagMAX *mir*Vana Total RNA Isolation Kit	96 preps	A27828
Thermo Scientific accessories		
KingFisher Flex Purification System with 96 Deep-Well Head	1 system	5400630
Applied Biosystems accessories		
Veriti 96-Well Thermal Cycler	1 instrument	4375786
ProFlex 96-Well PCR System	1 instrument	4484075
MicroAmp EnduraPlate Optical 96-Well Clear Reaction Plates with Barcode	20 plates	4483354
MicroAmp Optical 96-Well Reaction Plate	10 plates	N8010560
MicroAmp Clear Adhesive Film	100 films	4306311
MicroAmp 8-Tube Strip with Attached Domed Caps, 0.2 mL	125 strips	A30589

H = human, M = mouse, R = rat.
* Not all kits are available in all countries.

Find out more at **thermofisher.com/collibri**

ThermoFisher SCIENTIFIC