Seminars in Cell & Developmental Biology xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Birth, coming of age and death: The intriguing life of long noncoding RNAs

Samudyata[a], Gonçalo Castelo-Branco[a], Alessandro Bonetti[a,b,*]

[a] Laboratory of Molecular Neurobiology, Department Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[b] Division of Genomic Technologies, RIKEN Center for Life Science Technologies, RIKEN Yokohama, Japan

ARTICLE INFO

Article history:
Received 19 August 2017
Received in revised form 6 November 2017
Accepted 9 November 2017
Available online xxx

Keywords:
lncRNA
circRNA
RNA processing
eRNA
Transcription
Splicing
Polyadenylation
Subcellular localization

ABSTRACT

Mammalian genomes are pervasively transcribed, with long noncoding RNAs being the most abundant fraction. Recent studies have highlighted the central role played by these transcripts in several physiological and pathological processes. Despite several metabolic features shared between coding and noncoding transcripts, these two classes of RNAs exhibit multiple differences regarding their biogenesis and processing. Here we review such distinctions, focusing on the unique features of specific long noncoding RNAs.

Contents

* Corresponding author at: Laboratory of Molecular Neurobiology, Department Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.
E-mail address: alessandro.bonetti@ki.se (A. Bonetti).

https://doi.org/10.1016/j.semcdb.2017.11.012
1084-9521/

1. Introduction

Recent advances in genomic and transcriptomic fields have shown that the fraction of genomes transcribed into RNA without apparent protein-coding potential is larger than expected [1–3]. The vast majority of mammalian genomes are pervasively transcribed, accounting for a previously unappreciated complexity of the noncoding RNA (ncRNA) fraction [4]. Based on their length, ncRNAs can be further classified as either small ncRNAs or long noncoding RNAs (lncRNAs), the latter class being longer than 200 nucleotides (nt) [3]. Although structural ncRNAs such as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) have been studied for many decades, in the past few years lncRNAs have received growing attention for their functional importance [5]. Here, we will review the general features of lncRNA biogenesis and processing and subsequently we will focus on specific classes of lncRNAs that exhibit deviations from the general rules. While the life cycle of lncRNAs present several similarities to mRNAs, several subtle differences have been identified in the recent years. As technologies improve and other methodologies are established, unveiling and characterization of further differences might allow a more refined classification for lncRNAs.

2. Heterogeneity in lncRNA classification

Two main pre-requisites are needed for acceptance into the lncRNA "club": the first is a length of the RNA greater than 200 nt, while the second is the lack of any coding potential, i.e. no open-reading frame longer than 30 amino acids [6]. These selection criteria were established for convenience or technical reasons, rather than a strict length threshold determining functional or structural differences. The length cut-off was introduced due to the properties of the columns routinely used to purify RNAs from cell extracts, which retain RNAs longer than 200 nt while discarding shorter RNAs. Incidentally, this cut-off conveniently separates lncRNAs from other shorter ncRNAs such as microRNAs, Piwi-interacting RNAs, small nucleolar RNAs (snoRNAs), tRNAs, among others. In addition, it is worth mentioning that some extensively studied lncRNAs such as *XIST* and *H19* have open reading frames longer than 100 amino acids [7].

The lncRNA repertoire comprises thousands of transcripts, which have distinct biogenesis, subcellular localizations and molecular functions. Given such heterogeneity, the subdivision in different classes has proven to be a daunting task and so far there has been no comprehensive categorization. Most commonly, lncRNAs are classified based on their genomic location and close proximity to protein-coding genes (Fig. 1) [8]. Other more biologically relevant classifications have also been employed, based on subcellular localization (nuclear, cytoplasmic or equally present in both compartments) or mechanism of action on the chromatin (*cis*- or *trans*-acting depending on target genomic locus) [9]. As the function and structure of lncRNAs continue to be characterized and novel aspects are uncovered, general functional principles and structural domains are likely to be established. Hence, we foresee that scientific community will be able in the near future to define taxonomy with similar level of detail we currently have for proteins.

3. General features of lncRNA metabolism

Most landmarks in the metabolism of lncRNAs overlap with those of messenger RNAs (mRNAs). When examining several stages for RNA biogenesis of human coding and noncoding genes, Mukherjee and colleagues have identified seven classes with similar RNA metabolism profiles, including both mRNAs and lncRNAs to varying degrees [10]. Without adopting existing annotations, the authors found that within each class, coding and noncoding transcripts tend to be more similar than expected, further blurring the boundaries between mRNAs and lncRNAs. In this section, we will revisit the different stages of maturation of lncRNAs ranging from expression to degradation (Fig. 2). Where applicable, we will also point out differences with protein-coding transcripts, highlighting processes that are specific for noncoding transcripts.

3.1. Birth

3.1.1. Chromatin marks

The transcriptional regulation of lncRNAs and mRNAs present several similarities. Several studies have reported evolutionary conservation between lncRNA and mRNA promoters in human and mouse species [2,11,12]. Similar to mRNAs, promoters of lncRNAs present typical enrichment for active chromatin marks such as histone 3 lysine 4 trimethylation (H3K4me3), histone 3 lysine 9 acetylation (H3K9ac) and histone 3 lysine 27 monomethylation (H3K27me1) [11]. In a genetic screen to dissect determinants of bidirectional transcription, divergent antisense expression was enriched for histone 3 lysine 56 monoacetylation and repressed by the CAF-I (Chromatin Assembly Factor 1) pathway [13]. Divergent transcription from promoters transcribed by RNA polymerase II (pol II) is found in multiple eukaryotic organisms [14], with a subset hypothesized to be lncRNAs transcribed in antisense direction from promoters of active protein-coding genes [15]. A recent study from Melé and colleagues revealed enrichment for histone 3 lysine 9 trimethylation (H3K9me3) at promoters of active long intergenic noncoding RNAs (lincRNAs), a subclass of lncRNAs transcribed from DNA regions located between protein-coding genes [16]. In addition, although this histone modification is usually associated with repression of transcription, lincRNAs with higher H3K9me3 levels exhibit a more prominent tissue-specific expression [16]. These are the first examples of a specific chromatin modification that selective marks the transcription of lncRNAs. Identification of additional marks with similar properties will be instrumental not only to discern potential differences in the transcriptional mechanisms regulating coding and noncoding transcripts, but also to *de novo* identify lncRNAs.

3.1.2. Transcription initiation and 5′ capping

Similar to protein-coding genes, lncRNAs are mostly transcribed by RNA pol II, in contrast to many small RNA subclasses, which are transcribed with RNA pol III. Nevertheless, several noncoding repeated elements, such as Alu sequences in humans and short interspersed nuclear elements (SINE) B1 and B2 in mice, are transcribed by RNA polymerase III and can regulate the activity of RNA pol II in response to stress [17].

The recruitment of transcription factors (TFs) at transcription start sites (TSSs) to form the pre-initiation complex together with RNA pol II is very similar between mRNAs and lncRNAs, although the extent of activation may vary between classes. Notably, some

Please cite this article in press as: Samudyata, et al., Birth, coming of age and death: The intriguing life of long noncoding RNAs, Semin Cell Dev Biol (2017), https://doi.org/10.1016/j.semcdb.2017.11.012