Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

lncRNA in worms – Time to meet the neighbors

Soumasree De[1], Liron Levin[2] and Barak Rotblat[1]

Abstract

Long noncoding RNA or lncRNA have been in the spotlight in recent years, due to their abundance in the human genome and the crucial regulatory roles they play in diverse biological processes. The mode of action by which lncRNA regulate biological processes is not fully understood however, in many cases it was found that they function by regulating the expression of neighboring genes in *cis*. To date, lncRNA have received little attention from the *Caenorhabditis elegans* research community. In particular, lncRNA functions in *cis* have not been thoroughly investigated. Here we review the known functions of lncRNA in mammals and *C. elegans*. To promote lncRNA studies in *C. elegans*, we provide a catalog of lncRNA neighbor genes in *C. elegans*, their human homologs and the biological categories they belong to. We propose that *C. elegans* could be a powerful model for studying these enigmatic non-protein coding genes.

Addresses

[1] Ben Gurion University of the Negev, Israel

[2] Bioinformatics Core Facility, National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Corresponding author: Rotblat, Barak (rotblat@bgu.ac.il)

Current Opinion in Systems Biology 2018, ■:1–6

This review comes from a themed issue on **Systems biology of model organisms (2019)**

Edited by **Denise Kirschner** and **Ramit Mehr**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.coisb.2018.08.007

Keywords

Bioinformatics, Evolution, Transcription regulation, Synteny.

Introduction

LncRNA are a large and diverse class of transcripts ranging from 200 to 100,000 bp in length which do not code for proteins. LncRNA have been identified in plants, invertebrates and mammals [1]. Most lncRNA gene structure resembles protein coding mRNAs with regards to promotor CpG islands, RNA Poly II transcription, a multiexonic structure and, in some cases the transcript is decorated with a 3′ poly A tail and a 5′ m7GTP cap [2,3]. Although particular transcripts depicted as lncRNA were found to encode short open reading frames [4–6], in most cases there is no evidence that they do code for peptides [7]. LncRNA are classified by their location relative to the nearby protein coding genes, as long intervening non coding RNA (the genomic loci does not overlap the exons of protein coding genes), antisense lncRNAs (the loci present in the opposite strand of protein coding genes) and intergenic lncRNAs (the loci present between two protein coding genes) [8]. Genome based studies have revealed several functional roles for lncRNA in chromatin structure regulation, gene transcription, development and disease [9].

Evolution of lncRNA may have promoted complexity

LncRNAs have evolved from either partial or total duplication and divergence of consequent sequences [10,11]. They are linked to a wide range of developmental processes and are found in a diverse range of species, from sponges to primates [10]. A recent systemic investigation in demosponge has identified 2935 lncRNAs, most of them overlapping with protein coding genes [12]. Similarly, a large-scale evolutionary study of lncRNA repertoires and expression patterns in 11 tetrapod species has identified almost 11,000 primate-specific lncRNAs, some of which originated more than 300 million years ago [13].

An interesting possibility regarding lncRNA functions is that by fine tuning gene expression, lncRNA facilitated the increase in complexity found in the late stages of mammalian evolution and in humans in particular [14]. This notion is supported by the finding that there are a significant number of lncRNA associated with human complex traits and, importantly, these lncRNA are suspected of regulating the expression of neighboring protein-coding genes with significant enrichment for chromatin regulators [15]. From a systems biology point of view, one of the major challenges of the lncRNA field is to infer lncRNA function from their sequence [16]. Of note, recent efforts promoted our understanding of the information embedded in lncRNA sequences enabling prediction of their localization [17], protein binding [18] and binding to small RNA [19].

Mammalian lncRNA function to regulate chromatin

The functions of lncRNA in chromatin regulation have been the focus of lncRNA research from the inception of the field [20,21] with *XIST*, a lncRNA which promotes X inactivation maintenance by recruiting chromatin modifier proteins to the X chromosome [22,23], serving as a classic example for such lncRNA functions in *cis* [10].

Other lncRNA have been shown to play a role in regulation of chromatin states in *trans* by binding chromatin regulating proteins and targeting them to particular regions of the genome [24–28]. *HOTAIR*, yet another well studied lncRNA, was shown to be expressed from the human and mouse *HOXC* cluster and promote silencing of genes within the *HOXD* cluster by recruiting and targeting chromatin modifying enzymes to the *HOXD* locus, thus functioning in *trans* [29,30]. Recent analysis of the *HOTAIR* sequence, chromatin state and expression pattern revealed that it may function both in *trans* and *cis* to regulate gene expression within the HOX clusters [31]. These findings have raised the attractive hypothesis that lncRNA target chromatin regulator proteins to specific regions of the genome to regulate chromatin structure and thus regulate gene expression. Whether these lncRNA chromatin regulatory functions in *trans* are mediated by specific sequences within the lncRNA is still debated. One major chromatin regulating protein, PRC2, was shown to bind promiscuously to RNA [32,33] supporting an alternative model where the genomic location of the lncRNA, rather than the specific sequence, is the major determinant of its functions [33].

lncRNA regulate neighboring genes in *cis*

Progress has been made in better understanding the functions of lncRNA as regulators of adjacent genes in *cis* [34]. *Upperhand* is a lncRNA gene neighbor of the transcription factor *Hand2* coding gene which is essential for heart development. To test if *Upperhand* regulates *Hand2* in *cis*, Anderson et al. used TALEN to introduce either a transcription terminator or a fluorescent protein coding sequence into the second exon of the *Upperhand* gene and found that transcription termination, but not sequence swapping, eventuated in reduced *Hand2* expression and defective heart development in mice [35], supporting the model where the lncRNA functions in a sequence independent manner to activate an adjacent gene in *cis*. Similarly, this principal was demonstrated by using CRISPR to insert transcription termination sequences into mouse lncRNA and protein coding genes and analyzing the expression of the adjacent genes [36]. In both cases, inhibiting the transcription of a gene often resulted in reduced expression of the neighboring genes.

To functionally interrogate the role of 10,000 lncRNA in melanoma tumor cells drug resistance, Joung et al. used CRISPR activation and a pooled screen by which they identified 11 lncRNA promoting resistance [34]. In this study, most identified functional lncRNA promoted drug resistance by regulating one, or more, adjacent genes. Similarly, Bester et al. performed a CRISPR activation screen to identify both protein-coding and lncRNA genes promoting drug resistance in acute myeloid leukemia tumor cells and found that the lncRNA *GAS6-AS2* promotes resistance by activating the adjacent *GAS6* protein coding gene [37]. Interestingly, in a guilt-by-association analysis, lncRNA protecting tumor cells from chemotherapy were enriched for fatty acid metabolism and oxidative phosphorylation categories, suggesting that lncRNA are also involved in regulation of metabolism related genes [37]. Indeed, an earlier study showed that lncRNA transcription negatively regulates the adjacent *DHFR* gene activity [38].

How lncRNA promote or inhibit the expression of neighboring genes is still not completely clear [39]. Using computational and experimental analysis, Lou et al. investigated the functions of divergent lncRNA-protein coding gene pairs (as illustrated in the graphical abstract) which represent 20% of human lncRNA [14], and found that these lncRNA regulate the expression of adjacent genes, by a mechanism which includes binding of the lncRNA to the boarders of active chromosomal regions and to chromatin regulating proteins [39].

lncRNA in *Caenorhabditis elegans*

Despite the fact that many major discoveries in RNA biology have come from studying the small worm *Caenorhabditis elegans* (*C. elegans*) [40,41] less in known about lncRNA in this model organism. Nam et al. identified 170 lncRNAs in *C. elegans* [42], of which 25% have sequence conservation with endogenous micro RNA, and theoretically, could serve as templates for these small RNA species. The other lncRNA were more conserved and exhibited transcription patterns implicating them in developmental process and differentiation of worms. These lncRNA transcriptional patterns are linked to male identity, sperm formation, and interaction with sperm-specific mRNAs [42].

A recent study attempted to infer biological function of lncRNA in *C. elegans* by following their spatiotemporal expression pattern [43]. Using a GFP knock-in approach, the spatial expression patterns of 68 lncRNAs were examined in 18 tissue categories throughout eight developmental stages. This study found that lncRNA and miRNA promoters are less active at the embryonic stage, as compared with the promotors of transcription factors (TFs), but become comparable to TFs after embryogenesis. Finally, the expression pattern of this lncRNA gene set is similar to that of miRNAs and TFs in mature animals [43].

Several lncRNAs were also identified in the muscle transcriptome profile of dauer and aging worms [44]. Using tissue-specific RNA-seq, splicing-based RNA tagging (SRT), the authors detected 461 novel RNA transcripts in worm muscles, most of which are predicted to be lncRNA. Using reporter assay, 5 out of 8 tested transcripts were shown to be expressed in muscle. It is