Published in final edited form as:

Trends Genet. 2017 October ; 33(10): 660–662. doi:10.1016/j.tig.2017.07.005.

On the origin of lncRNAs: missing link found

Joaquín M. Espinosa^{1,2}

¹Linda Crnic Institute for Down Syndrome and Department of Pharmacology, University of Colorado Anschutz Medical Campus, Aurora, Colorado 80045, USA

²Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Boulder, Colorado 80203, USA

Abstract

Non-coding RNAs known as enhancer-derived RNAs (eRNAs) and long non-coding RNAs (lncRNAs) have received much attention, but their true functional specialization and evolutionary origins remain obscure. The recent characterization of Bloodline, an eRNA derived from a super-enhancer that also functions as a lncRNA, suggests that lncRNAs can evolve from eRNAs.

Keywords

lncRNA; eRNA; super-enhancer; non-coding RNA; chromatin loop; CTCF; topological activation domain

One of the most captivating observations of the genomics era is that the fraction of genomes that is transcribed into RNAs without protein coding capacity is much larger than anticipated [1]. Much research has been devoted to defining the biological roles and mechanisms of action of two classes of non-coding RNAs (ncRNAs), those known as enhancer-derived RNAs (eRNAs) and those known as long non-coding RNAs (lncRNAs). eRNAs have been characterized as short, unspliced, non-polyadenylated RNAs that are transcribed bidirectionally from enhancer DNA elements [1]. eRNAs are unstable, and often accumulate at the site of transcription where they can stabilize enhancer-promoter interactions and transcription factor (TF) binding [1] (Figure 1A). In contrast, lncRNAs are transcribed from their own promoter, are longer, and usually spliced and polyadenylated [1]. lncRNAs are more stable and abundant than eRNAs, and some have been shown to perform functions in *trans* (Figure 1B). Despite these seemingly strong discriminating criteria, many recent reports have brought into question the distinction between eRNAs and lncRNAs. In some instances, the bulk of the biological activity of lncRNA loci was found to be due to an enhancer DNA element near the lncRNA transcription start site [2, 3]. In other cases, lncRNA loci were shown to actually work locally in an RNA-independent fashion to facilitate chromatin accessibility or regulate enhancer action by the mere act of transcription

Correspondence: joaquin.espinosa@ucdenver.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

[4, 5]. Examples of such ‘lncRNA-to-eRNA converts’ are linc-p21, now considered an eRNA derived from a p53 enhancer element in the human *CDKN1A* (p21) locus [2], and Lockd, an RNA derived from the human *CDKN1B* (p27) locus [3]. Based on these observations, I recently hypothesized an evolutionary path from eRNAs to lncRNAs [6], whereby spurious splicing and polyadenylation of ancestral eRNAs, due to the accidental presence of splicing and 3′ processing sequences in the vicinity of the respective enhancers, would lead to the stabilization of the eRNAs. In turn, these stable and abundant eRNAs would provide the fodder for evolutionary selection of novel biochemical functions in *trans* via fine tuning of incipient RNA-RNA and RNA-protein interactions (Figure 1B).

Perhaps nothing in the study of evolution has converted more skeptics than the discovery of transitional fossils, those ancient species whose mere anatomy captures obvious evolutionary transitions. Even the staunchest creationists should bow to the Tiktaalik, the extinct transitional four-legged fish loaded with mosaic features indicative of its explorations out of the aqueous, into the dry. Now, Alvarez-Dominguez and colleagues [7] report the identification of the RNA equivalent of the *Periophthalmus barbarous* mudskipper, a contemporary fish that can walk on land for days and eventually return to water, the living version of Tiktaalik; that is, an eRNA that is also a lncRNA. Their interesting findings not only reinforce the need for a revision of the eRNA/lncRNA divide, but, as explained below, they also illuminate a novel mechanism for coordinated gene expression that overrides cell-type specific chromatin landscapes.

Super-enhancers, also known as stretch enhancers (SEs), are DNA sequences harboring large clusters of TF binding sites near genes whose activity is required for cell type-specific processes [8]. SEs encompass chromatin domains that are actively transcribed to produce large amounts of eRNAs, and a few SEs often account for more than half of the total chromatin marks associated with enhancer activity in a given cell type. While cataloging SEs in erythroblasts, Alvarez-Dominguez et al found that one of the strongest SEs was located upstream of the *SLC4A1/BAND3* gene encoding an erythrocyte-specific membrane transporter [7]. Interestingly, this SE, which produced the most abundant eRNA, overlaps a genomic region previously shown to encode the lncRNA known as alncRNA-EC7/Bloodlinc [9]. Using a series of mechanistic investigations including genome-editing, the authors demonstrated that the promoter region of the *Bloodlinc* locus functions as a potent enhancer to drive *SLC4A1* expression. Consequently, Bloodlinc should be classified as an eRNA. However, when they depleted Bloodlinc RNA using shRNA technology, without perturbing the underlying DNA sequence, Bloodlinc was found to be required for *SLC4A1* expression, as well as expression of many other erythrocyte-specific genes located elsewhere in the genome. Additional experiments demonstrated that the Bloodlinc RNA escapes its site of transcription to bind hundreds of sites across the genome, often near genes whose expression is affected by Bloodlinc depletion. Mechanistically, the Bloodlinc RNA was found to bind several proteins, most prominently HNRNPU, a RNA-binding protein with chromatin regulatory properties, whose depletion partially mimicked the effects of Bloodlinc inactivation. Perhaps the most striking result in this study is that Bloodlinc depletion impairs erythroblast differentiation, while its overexpression drives differentiation into erythrocytes, something that is not reproduced by manipulation of *SLC4A1* expression. Altogether, these

results indicate that Bloodline represents a missing link in the eRNA to lncRNA evolutionary path.

The discovery of Bloodline forces us to revise our understanding of enhancer function and evolution. According to the current consensus, enhancer action is largely restricted to within topological activation domains (TADs), large chromatin regions within which molecular interactions are favored, and which boundaries are defined by chromatin insulating factors such as CTCF and cohesins [10] (Figure 1A). In this model, an enhancer can regulate only genes within the TAD, because intervening insulators block enhancer-promoter communication. Since TAD architecture varies from cell type to cell type, TADs are considered to be key determinants of lineage-specific gene expression [10]. However, the discovery of Bloodline indicates that some enhancers can also regulate the expression of genes located in distant TADs, and that they do so *indirectly*, via production of eRNAs that act as lncRNAs with *trans* functions (Figure 1B). Thus, as a typical enhancer evolves into an SE to drive the robust expression of a key cell type-specific gene, evolutionary forces can co-opt the abundant eRNA produced to ensure coordinated expression of other genes in the genome required for cell type-specific transcription. In this view, cell type-specific TFs and eRNA/lncRNAs work sequentially to elicit a transcriptional cascade to drive cells into a differentiated state (Figure 1B).

Nothing in nature happens only once. Not wings, not fins, not eyes. How many SEs produce RNAs that work both as *cis*-acting eRNAs and *trans*-acting lncRNAs? Bloodline is unlikely to be the only living fossil. To my fellow molecular biologists I ask: have you checked whether your super-enhancer RNA has become a lncRNA?

References

1. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 2014; 157:77–94. [PubMed: 24679528]
2. Groff AF, et al. In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements. *Cell reports*. 2016; 16:2178–2186. [PubMed: 27524623]
3. Paralkar VR, et al. Unlinking an lncRNA from Its Associated cis Element. *Mol Cell*. 2016; 62:104–110. [PubMed: 27041223]
4. Schultz BM, et al. Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. *Nucleic Acids Res*. 2015; 43:745–759. [PubMed: 25539921]
5. Engreitz JM, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*. 2016; 539:452–455. [PubMed: 27783602]
6. Espinosa JM. Revisiting lncRNAs: How Do You Know Yours Is Not an eRNA? *Mol Cell*. 2016; 62:1–2. [PubMed: 27058782]
7. Alvarez-Dominguez JR, et al. The Super-Enhancer-Derived alncRNA-EC7/Bloodline Potentiates Red Blood Cell Development in trans. *Cell reports*. 2017; 19:2503–2514. [PubMed: 28636939]
8. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013; 153:307–319. [PubMed: 23582322]
9. Alvarez-Dominguez JR, et al. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood*. 2014; 123:570–581. [PubMed: 24200680]
10. Matharu N, Ahituv N. Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS genetics*. 2015; 11:e1005640. [PubMed: 26632825]

Figure 1. Evolution of lncRNAs

(A) Typical enhancers are transcribed bidirectionally to produce eRNAs, which are short, unspliced, non-polyadenylated RNAs that can work in *cis* to stabilize enhancer-promoter communication and transcription factor (TF) binding. Insulator elements bound by CTCF and cohesins usually restrain the action of a typical enhancer. Topological activation domains (TADs) are large chromatin domains encompassing many genes and delimited by strong insulator elements. (B) Super-enhancers evolved to ensure robust expression of cell type-specific genes via selection of clusters of TF binding sites. Consequently, transcription over super-enhancers covers a longer distance, which increases the chances of spurious splicing, polyadenylation and stabilization of the RNA. Evolutionary forces can co-opt these stable RNAs to confer them with *trans*-functions at loci outside of the TAD of origin via fine-tuning of accidental RNA-RNA or RNA-protein interactions. In this scenario, distal

regulation by a lncRNA derived from a super-enhancer would further reinforce cell type-specific transcription and cellular identity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript