Molecular Cell
Previews

Not Just Noise: Genomics and Genetics Bring Long Noncoding RNAs into Focus

L. Stirling Churchman[1,*]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: churchman@genetics.med.harvard.edu
http://dx.doi.org/10.1016/j.molcel.2016.12.017

Thousands of long noncoding RNAs (lncRNAs) have been annotated, yet their functions remain unclear. Recent studies—including Schlackow et al. (2017) in this issue of *Molecular Cell*—using orthogonal methods investigated the expression and functions of lncRNAs, resulting in deeper appreciation for the salient differences between lncRNAs and mRNAs and the roles lncRNAs serve.

Tiling microarrays and next-generation sequencing technologies revealed that the genome is pervasively transcribed, initiating a new era of RNA biology. Many transcription events outside of coding genes generate long noncoding RNAs (lncRNAs), which, as their name suggests, are longer than 200 nt and do not produce protein products. Several lncRNAs, such as Xist, HOTAIR, and MALAT1, have well-characterized functions related to their RNA sequences (Quinn and Chang, 2016). On the other hand, many lncRNAs are poorly conserved and expressed at low levels, suggesting that lncRNAs may play their functional roles in the aggregate, rather than individually, or that the act of transcription of some lncRNAs (rather than the transcripts themselves) imparts function—or, alternatively, that many lncRNAs simply reflect transcriptional noise (Quinn and Chang, 2016). In an effort to distinguish between these possibilities, several recent studies have exploited genome editing tools, ENCODE datasets, and nascent RNA profiling to characterize the production and function of lncRNAs (Engreitz et al., 2016; Liu et al., 2016; Melé et al., 2016; Schlackow et al., 2017). These efforts identified sets of lncRNAs that are likely to exert specific functions, enabling researchers to interrogate individual lncRNAs within these sets with greater confidence of obtaining biological insight. In addition, the studies revealed new attributes of lncRNAs and established new experimental strategies for dissecting their expression and function.

To investigate whether and how lncRNAs differ from mRNAs, in separate studies, the Rinn and the Proudfoot laboratories analyzed how lncRNAs are created, processed, and degraded (Melé et al., 2016; Schlackow et al., 2017). The Rinn group used ENCODE data to assess the differential enrichment of histone modifications in the promoter regions of lncRNAs and mRNAs and found modest, but significant, differences between the two types of transcripts. Consistent with previous reports, both studies found that splicing of lncRNAs was not nearly as efficient as mRNA splicing. In fact, the Rinn group found that splicing efficiency was the predominant difference in the production of the two classes of transcripts.

In this issue of *Molecular Cell*, to determine how lncRNAs are transcribed, Schlackow et al. (2017) used a variant of the native elongating transcript sequencing (NET-seq) approach, which maps RNA polymerase (Pol) II density genome wide at nucleotide resolution (Mayer et al., 2015; Nojima et al., 2015). Using phospho-specific antibodies, Schlackow et al. (2017) mapped the enrichment of five Pol II C-terminal domain (CTD) phosphorylation marks (Nojima et al., 2015; Schlackow et al., 2017). The Pol II CTD is an unstructured region consisting of 52 heptapeptide repeats ($Y_1S_2P_3T_4S_5P_6S_7$), which are dynamically phosphorylated during transcription at each of the serine, threonine, and tyrosine residues and recruit regulatory factors during key stages of elongation, such as the transition to productive elongation, splicing, and termination (Harlen et al., 2016; Heidemann et al., 2013). Schlackow et al. (2017) observed that threonine-4 phosphorylation (Thr4P) was strongly enriched across entire lncRNA transcription units. By contrast, in mRNAs, Thr4P was sharply enriched after polyadenylation sites. Thr4P is a mark associated with transcription termination; it is enriched in regions of termination and the mark interacts with a key termination factor in *S. cerevisiae* (Harlen et al., 2016; Heidemann et al., 2013). Thus, these data suggest that the regions in which lncRNA transcripts are terminated are much broader than the corresponding regions of mRNAs.

Prematurely terminated transcripts are often targeted to the nuclear exosome for rapid degradation. Schlackow et al. (2017) found that depletion of a nuclear exosome component resulted in widespread stabilization of lncRNAs, leading them to propose that many lncRNAs are co-transcriptionally targeted for nuclear degradation. However, when Melé et al. (2016) investigated the stability of lncRNAs by an orthogonal approach using the transcription inhibitor actinomycin D, they found that lncRNAs and mRNAs had comparable lifespans. These two analyses reveal that lncRNAs have two alternative fates: (1) rapid degradation within the nucleus by the nuclear exosome, a fraction revealed by exosome depletion, or (2) stabilization, a fraction revealed by transcription inhibition. In other words, many lncRNAs are degraded immediately, but a sizable fraction escape this fate and accumulate.

Both groups identified a subset of lncRNAs that are transcribed and processed in a manner similar to mRNAs and proposed that these RNAs are more likely to have physiological functions. Two other recent studies that used genetic approaches to investigate lncRNA function provide models for future functional analyses. The Lim and Weissman

laboratories performed a CRISPR interference (CRISPRi) screen to query the function of 16,401 lncRNAs in seven cell lines (Liu et al., 2016). CRISPRi uses a nuclease-dead Cas9-KRAB enzyme and a single guide RNA to repress transcription across a targeted 1 kb window of the genome. Using growth rate as a phenotype, Liu et al. (2016) found that a small percentage (3%–8%) of lncRNAs impacted growth in any one cell line and that those involved in regulating growth did so in a highly cell-type-specific manner. In a more focused effort, the Lander laboratory used genetic editing to investigate the functions of 12 lncRNA loci and found that five of them acted by modulating the expression of a nearby gene (within 10 Mb) (Engreitz et al., 2016). In all five cases, the RNA itself was not required; instead, the act of transcription or splicing was essential for the effect. In four of these cases, the lncRNA promoter was essential for the expression of a nearby gene, i.e., the lncRNA promoter was acting as an enhancer, and the lncRNA represents a relatively long and stable enhancer RNA.

Together, these studies and others suggest that there are two classes of functional lncRNAs: (1) those whose RNA is required for function and (2) those whose expression and processing confer function. These classes are in line with a model of de novo gene formation proposed by the Sharp laboratory (Wu and Sharp, 2013), in which noncoding divergent transcription arising from promoters of coding genes and distal enhancers serve as a source of RNAs that can evolve into new protein-coding genes. In support of their model, Wu and Sharp (2013) described a promoter of a hominoid-specific protein-coding gene, *MYEOV*, that corresponds to a region of the mouse genome that bears chromatin marks associated with enhancers. Wu and Sharp (2013) proposed that lncRNAs represent intermediates in this evolutionary trajectory, having evolved from enhancer RNAs (eRNAs) after acquiring splicing-related *cis* elements. Consistent with this, the study by the Lander lab describes 11 examples of mouse lncRNAs whose promoters correspond to conserved DNA sequences that have enhancer chromatin signatures in human cells (Engreitz et al., 2016), suggesting that the mouse lncRNAs might be derived from eRNAs. Therefore, in addition to the two classes of functional lncRNAs, there may exist a third class of lncRNAs that are derived from eRNAs and serve as a reservoir of RNAs that can subsequently be shaped by evolutionary pressures and serve physiological functions in descendants. This possibility raises several key questions. How do these transcripts arise in the first place? Where do they come from? Do enhancers naturally produce bidirectional transcription from the onset? To answer these basic and foundational questions, future studies should seek to elucidate the intrinsic nature and evolution of promoter/enhancer transcription.

REFERENCES

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Nature *539*, 452–455.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Cell Rep. *15*, 2147–2158.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Biochim. Biophys. Acta *1829*, 55–62.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2016). Science. Published online December 15, 2016. http://dx.doi.org/10.1126/science.aah7111.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Cell *161*, 541–554.

Melé, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C., and Rinn, J.L. (2016). Genome Res. Published online December 7, 2016. http://dx.doi.org/10.1101/gr.214205.116.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Cell *161*, 526–540.

Quinn, J.J., and Chang, H.Y. (2016). Nat. Rev. Genet. *17*, 47–62.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Mol. Cell *65*, this issue, 25–38.

Wu, X., and Sharp, P.A. (2013). Cell *155*, 990–996.