
Twenty years

SIDNEY ALTMAN

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, USA

The most significant event of the past 20 years in molecular biology is the discovery of DNA sequences that are transcribed into RNAs that do not code for proteins. This phenomenon, originally found in *C. elegans*, has been extended to many organisms, including *H. sapiens*, and culminates to date in the further description of lnc RNAs (long non-coding RNAs). lnc RNAs are extremely long (>200 nts) and are transcribed from many parts of a genome, both from the Watson and Crick strands, and in humans, cover so many parts of the genome that the term “junk” DNA is rendered obsolete and incorrect. The future of RNA, for at least the next 10 years, is a most promising area of research that was founded by these recent discoveries.

The functions of the now partly characterized populations of these molecules are apparent when they determine phenotypes as the result of stress on normal housekeeping genes, and the same can also be true of short RNAs that do not code for proteins in bacteria. “Stress” can be defined as a cause of differentiation in eucaryotes and in procaryotes as adaptation to various external and regulatory factors.

Now the difficult part. Various states of differentiation in eucaryotes, which frequently includes complex genetic switches, are subject to the control of miRNAs. The multitude of actions of these small RNAs are more completely displayed when one reads the latest issues of the journals dealing with modern molecular genetics or the properties of RNA. The lnc RNAs also feature prominently as enhancers or repressors of gene function. They frequently involve action on complex arrangements of genes on one or both strands of DNA in a particular region of a chromosome. While only a few, perhaps two dozen or so, lnc RNAs have had their function identified, they offer sufficient evidence of the critical function these RNAs play in the growth and development of eukaryotic cells.

The ultimate questions are why these lnc RNAs are restricted to gene function in differentiation and how did they evolve to be the masters of these events? No current answer is satisfactory. What is the trigger to begin transcription of these RNAs? We may know what physical-chemical external factors do but that is not enough. How did the expression of lnc RNAs begin, in a primordial sense? In fact, it seems obvi-

ous, that these RNAs might very well have been ancient vestiges of early transcription events. As time progressed, these long RNAs, because of their partially complementary binding properties to other RNAs, would get engaged in the functions that we now identify.

A simple idea, perhaps a too facile one, might account for the synthesis of lnc RNAs in all organisms. The *first* cell capable of replicating itself, and of constantly transcribing undifferentiated lnc RNAs on both strands of DNA by its single polymerase, would be the foundation of what we see today. All further replicating cells derived from the first one would also make lnc RNAs, which as evolution proceeds, would be divided into mRNAs from “genes” and miRNAs as described below. That single polymerase might also have replicated DNA and transcribed it, too, but we leave that question alone at this moment. We also leave alone an earlier non-enzymatic transcription in a non-replicating primitive bag of tricks. With this vestigial organism, or contemporary with it, evolution of the nucleic acids in response to mutations and other external properties that are engaged in food and metabolism will lead to the presence of genes, that is regions of the DNA that code specifically for peptides or proteins. The presence of “genes” or, alongside with this aspect of evolution, the number and size of the very long transcripts of RNA will decrease as “genes” arose. The lnc RNAs, through mutations, the presence of self-cleaving RNAs, and regions that are now susceptible to the action of ancient RNA enzymes will, as indicated, become shorter but by how much we do not know. The presence of nucleotide sequences that can bind to specific regions in DNA, or to newly made proteins, through the complementarity of sequences, will add to the increasing complexity of the RNA population and the function of “genes.” Much of the genome, especially in procaryotes, is used up by creation of new “genes” and the amount and size of nc RNAs is consequently decreased. We now have a bipartite system long, primordial RNA and “genes” and shorter RNA made from these genes.

The way in which genotypes and phenotypes today are affected by large decreases and increases through mutations in the genome, and therefore, in the size of lnc RNAs cannot be predicted. Such changes could be noticed in genes coding for

Corresponding author: sidney.altman@yale.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.049924.115>. Freely available online through the RNA Open Access option.

© 2015 Altman This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

proteins, but transcription that might occur on the opposite strand of these genes or in non-coding regions currently cannot easily be understood. However, it should be possible to engineer point mutations that are neutral in coding regions but do affect lnc RNAs on the opposite strand to determine if changing the sequence of these RNAs affects their function. Similarly, extensive deletions or additions of sequence will work in this way, too. In fact, in primates, additions of portions of lnc RNAs seemed to have already occurred and the newly evolved lnc RNAs are expressed primarily in a tissue specific manner. Their function is still not understood.

In prokaryotes, external shocks (pH, cold, lack of O₂; UV irradiation) select some nc RNAs that can bind to “genes” to amplify or decrease their expression. While these phenomena will lead in the future to systems we can explore today with modern molecular genetics, evolution triggered by physical and chemical aspects of the environment will again limit in terms of new mRNAs the number of nc RNAs with no function. Evolution provides a scheme for the integration of some fraction of nc RNAs in gene function. In *E. coli*, the amount of the genome not transcribed at all might be close to 10% or lower and the total amount of nc RNAs is not yet completely known.

In eucaryotes, the same ideas hold for the evolution of “genes.” Differentiation occurs after evolution selects for lnc RNAs that bind to effectors/repressors and that are active in evolution of “genes.” Similarly, miRNAs are used in the same way, but they have evolved much further in terms of dealing with a myriad of functions of particular genes involved in gene expression and differentiation. The amount of lnc RNAs transcribed from the genome is ~70% as determined by a scheme to annotate all RNAs in the transcriptome, but the amount with no function in humans has to be determined. *C. elegans* and yeast have a much smaller number of lnc RNAs. lnc RNAs that preserve their non-function with respect to regulating genes are still remnants of primordial RNA.

The number of miRNAs made from lnc RNAs in humans is about 6% of the total. However, miRNA in “genes” today and mRNAs are much further along in terms of specialized evolution in comparison with lnc RNAs. Certainly, the complete characterization of structure and function of, especially, the lnc RNAs will add to the knowledge base that might explain the original presence of these RNAs and what that means to today’s cells, both eukaryotic and prokaryotic in which these RNAs appear to be the masters of gene regulation.

What of the future? There is little confidence in making predictions about anything that might occur over the next 10 years, or even five years. The most fundamental questions regarding miRNAs and lnc RNAs as just outlined above may be answered. That is a wonderful thought and I might be willing to bet a quarter on that answer appearing, but not more than a quarter. I would bet more on the secondary possible answer to that question in which I shall paraphrase an answer that Sydney Brenner once gave on the future of molecular genetics in the early 1960s: All the details (*author’s explanation: trivia*) will be worked out by scientists from North America. But who will postulate the answer to the bigger question and supply the needed and simple experiments to give evidence for the answer?

There are two more points to be made briefly. Another engaging question for the future regards an understanding of brain function and the integration of neuron cellular output to form “thoughts” or “consciousness.” Surely, the study of non-coding RNAs will provide a complex and ultimately successful picture of their function in the brain. The time to sign up for that project is now.

Finally, will *RNA* survive as a front-line, important journal and continue to provide useful information to RNA researchers? To date, we can thank Tim Nilsen for the services he has provided as Editor-in-Chief, and we can only hope that the future of the journal will be as successful as it has been for the past 20 years.