

MINIREVIEW

Reprogramming of the non-coding transcriptome during brain development

Saba Valadkhan* and Timothy W Nilsen*

See research article <http://www.biomedcentral.com/1471-2202/11/14>

Abstract

A recent global analysis of gene expression during the differentiation of neuronal stem cells to neurons and oligodendrocytes indicates a complex pattern of changes in the expression of both protein-coding transcripts and long non-protein-coding RNAs.

A fascinating and unexpected outcome of the recent analyses of higher eukaryotic genomes has been the demonstration of pervasive transcription from non-protein-coding genomic sequences. Indeed, the preliminary results of the human ENCODE project indicate that whereas protein-coding sequences occupy less than 2% of the human genome, close to 93% of the genome is transcribed into RNA [1]. Although intronic sequences occupy a significant percentage of the non-protein-coding sequences in the genome, the majority of the independent non-protein-coding transcripts belong to the group of long non-coding RNAs (lncRNAs) - RNAs that are more than 200 nucleotides in length and do not appear to have any protein-coding potential [2-4]. A few members of this mysterious and highly understudied group of RNAs have been known for a long time, for example the *Xist* and *Air* RNAs; however, the majority of these transcripts have been only recently discovered in high-throughput transcriptome analyses. Furthermore, most of them are expressed at low levels and many do not show a high level of sequence conservation. Thus, the functional significance of this class of RNAs as a whole is still very poorly understood and subject to debate and speculation.

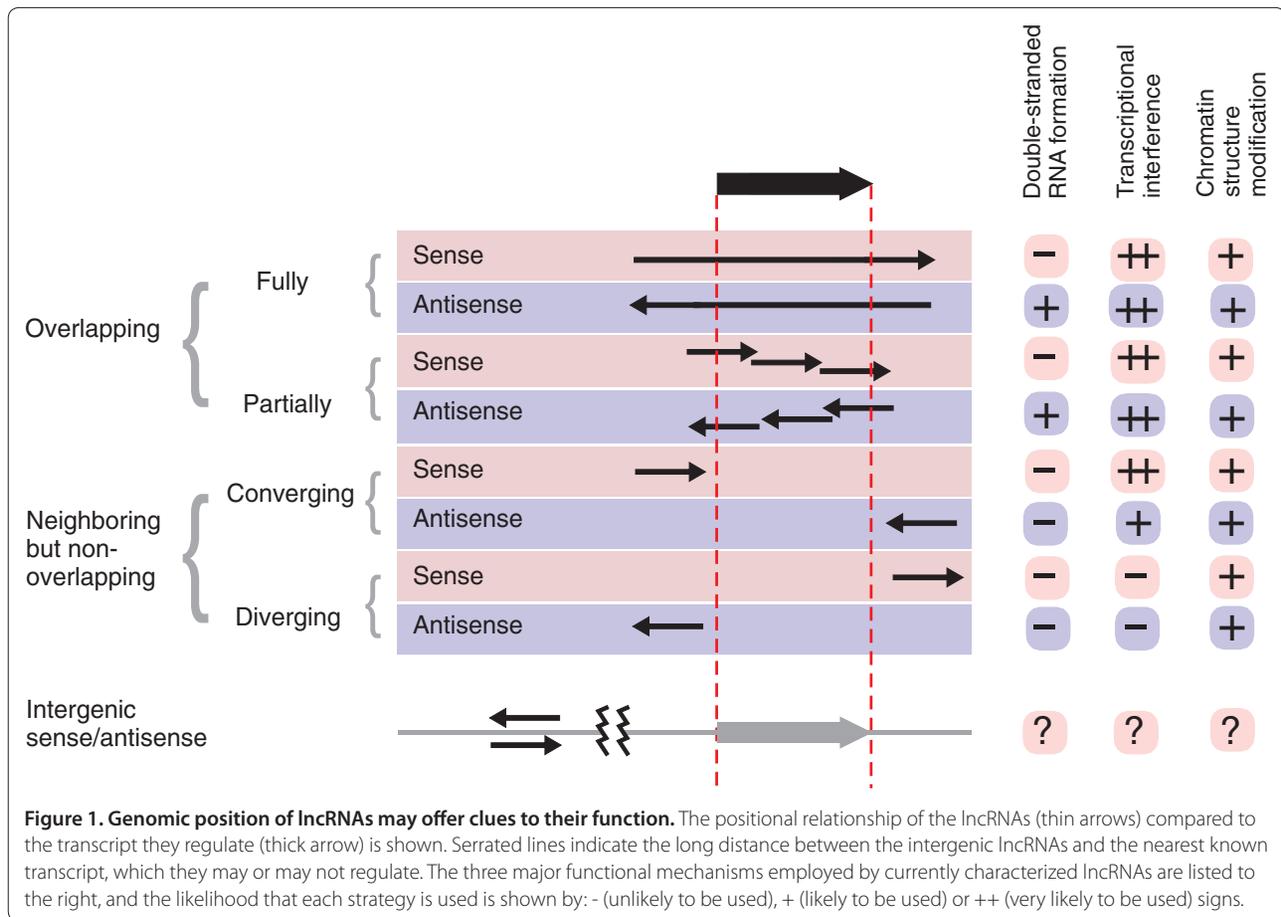
Large non-coding RNAs: a novel class of regulators or transcriptional noise?

Although our understanding of the biological role of this class of RNAs is rudimentary, there are several studies

that suggest that lncRNAs are much more than mere 'transcriptional noise,' or random output of background transcription, in higher eukaryotes. An interesting clue to the importance of this class of transcripts as a whole comes from the comparison of the percentage of the genome dedicated to non-coding sequences in organisms of differing complexity: as complexity increases so does the extent of non-protein-coding genomic sequences [5]. As the rather minor interspecies differences in the proteome cannot fully account for the dramatic increase in the level of complexity seen in higher eukaryotes, it is plausible that the non-coding transcriptome with its rapid rate of evolution may play a part in this process. Interestingly, bioinformatic analyses of the genomic regions that have evolved most rapidly between human and other primates point to several non-coding sequences, one of which is transcribed into a brain-specific long non-coding RNA that is expressed during the development of the human cortex [6]. Other studies have also indicated that a large fraction of the lncRNAs are expressed in brain, further supporting the tantalizing possibility that they might be involved in the development of the daunting complexity of the human brain [2-4].

Perhaps the most convincing evidence for a functional role for lncRNAs comes from studies that indicate that rather than resulting from background transcription, the expression of the non-coding transcripts is both temporally and spatially regulated. Several high-throughput analyses have shown tissue-specific expression of lncRNAs in stem cells, neuronal tissues and lymphocytes, among other tissues [7-10]. It has also been shown that stimulation of cultured macrophages with immunogenic stimuli results in the induction of the expression of a specific group of lncRNAs [7], proving that the expression of at least some of the lncRNAs is regulated. Interestingly, in many cases the tissue-specific lncRNA genes seem to be positioned in proximity to protein-coding genes with a known functional role in that tissue, suggesting the possibility of regulation *in cis* by these RNAs [2-4,7]. The above studies provide evidence for a functional role for at least a fraction of lncRNAs, but in order to determine the extent to which lncRNAs participate

*Correspondence: saba.valadkhan@case.edu; tw@case.edu
Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH 44106, USA



in cellular processes, more extensive and in-depth studies of the expression pattern of this group of transcripts and follow-up functional analyses are required.

In a recent in-depth global analysis of lncRNA expression, published in *BMC Neuroscience*, Mercer *et al.* [11] custom designed microarrays to analyze the changes in the expression pattern of both protein-coding transcripts and long non-coding RNAs in forebrain-derived mouse neural stem cells as they differentiate to GABAergic neurons and oligodendrocytes. Initial analysis of their results indicated that in parallel with up- and down-regulation of mRNA expression, the expression of a significant number of lncRNAs was also altered during neuronal and oligodendrocytic differentiation events. The expression of 16% of the approximately 14,800 interrogated protein-coding transcripts and 5% of the approximately 3,600 analyzed lncRNAs was significantly changed at one or more differentiation steps in these studies, with the altered expression of several members of both groups exclusively occurring during a single differentiation step. A number of previously characterized neuronally expressed lncRNAs were among those with altered expression, results that support the validity of the analyses.

Positional clues to the function of lncRNAs

While the above data suggest that the expression pattern of lncRNAs is at least as complex as that of mRNAs, a crucial question is the functional significance of the observed changes in lncRNA expression. To date, the molecular mechanism of function of the majority of lncRNAs remains unknown. However, the most informative clues to their possible mode of function come from their genomic position in relation to other transcripts. In many of the studied examples, lncRNAs have been found within transcriptionally complex loci where their expression, directly or indirectly, influences their neighboring genes [2-4]. An lncRNA may partially or completely overlap another gene in the sense or antisense direction, or it can be located in the close vicinity of another gene in the converging or diverging sense or antisense orientation without overlapping it (Figure 1). Depending on the exact position, the lncRNA transcript may affect the neighboring gene through formation of double-stranded RNA, or cause transcriptional interference or alter the local chromatin structure merely by being transcribed. There are also several known examples of intergenic lncRNAs, transcripts

that are located far away from other known transcripts and that are likely to exert their cellular function, if any, *in trans*, through mechanisms yet to be elucidated.

As a first step toward understanding the functional significance of the observed gene expression patterns, Mercer *et al.* [11] analyzed the genomic loci of the lncRNAs that showed significant changes in expression in their analysis. Interestingly, several of these lncRNAs, which included a number of novel transcripts, were part of transcribed loci that contained protein-coding genes with a known function in neural development. In many cases, the position of the lncRNA-mRNA pair was conserved between mouse and human, and expression analyses indicated coordinated expression, suggesting a functional interaction. A number of other lncRNAs were associated with highly conserved enhancer elements that regulate the development of forebrain, and yet another group overlapped brain-specific microRNAs (miRNAs), suggesting functional roles in the development of the nervous system for at least a subgroup of the lncRNAs studied. For example, a novel mRNA-like lncRNA that is both spliced and polyadenylated - AK044422 - overlaps a highly conserved and abundant brain-specific miRNA, miR-124a, and furthermore shows a complementary expression pattern with *ptbp1* (encoding polypyrimidine tract binding protein 1), a target of miR-124a. Analysis of the secondary structure of the lncRNA suggests that it might host several miRNAs, in addition to miR-124a, but whether it has a functional role beyond hosting miR-124a remains to be determined.

The intriguing associations observed by Mercer *et al.* [11] underscore several unanswered questions in the lncRNA field and at the same time provide a firm foundation for future in-depth studies aimed at addressing these questions. To what extent does the observed up- or downregulation of the analyzed lncRNAs affect lineage specification and differentiation of the neuronal stem cells? Do lncRNAs have master regulatory functions in the development of the central nervous system or indeed, in all developmental pathways, or are they confined to minor, fine-tuning regulatory roles? If so, what are the main strategies used by lncRNAs and can we predict their molecular mechanism of function by analysis of their sequence and genomic position? Further studies of mechanism and the dissection of the function of the lncRNAs will be partially guided by analysis of lncRNA secondary structure, as done in the present study. This indicated the presence of several conserved secondary structure elements that may correspond to hitherto unknown RNA functional domains.

Taken together, studies so far have provided us with a first glimpse of the intricate and complex web of interactions between lncRNAs and protein-coding RNAs and herald the emergence of a new paradigm for the developmental and differentiation processes of higher eukaryotes. While it is tempting to speculate that many existing gaps in our knowledge of cellular development and function may reflect a lack of knowledge of lncRNA functions, defining the extent to which this class of RNAs affects the development and function of higher eukaryotic species awaits detailed biochemical, molecular and cell biological analyses.

Published: 5 February 2010

Reference

1. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, *et al.*: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007, **447**:799-816.
2. Mercer TR, Dinger ME, Mattick JS: Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009, **10**:155-159.
3. Ponting CP, Oliver PL, Reik W: Evolution and functions of long noncoding RNAs. *Cell* 2009, **136**:629-641.
4. Wilusz JE, Sunwoo H, Spector DL: Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009, **23**:1494-1504.
5. Taft RJ, Pheasant M, Mattick JS: The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* 2007, **29**:288-299.
6. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, Kern AD, Dehay C, Igel H, Ares M Jr, Vanderhaeghen P, Haussler D: An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 2006, **443**:167-172.
7. Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM, Grimmond SM, Hume DA, Hayashizaki Y, Mattick JS: Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* 2006, **16**:11-19.
8. Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, Mattick JS: Genome-wide identification of long noncoding RNAs in CD8⁺ T cells. *J Immunol* 2009, **182**:7738-7748.
9. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS: Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 2008, **18**:1433-1445.
10. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS: Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* 2008, **105**:716-721.
11. Mercer TR, Qureshi IA, Gokhan S, Dinger ME, Li G, Mattick JS, Mehler MF: Long noncoding RNAs in neuronal-glia fate specification and oligodendrocyte lineage maturation. *BMC Neuroscience* 2010, **11**:14.

doi:10.1186/jbiol197

Cite this article as: Valadkhan and Nilsen: Reprogramming of the non-coding transcriptome during brain development. *Journal of Biology* 2010, **9**:5.