Long 'Non-Coding' RNAs: coding or non-coding? that is the question

Etienne Leygue[1], Jiuyong Xie[2]

[1]Department of Biochemistry & Medical Genetics,

[2]Department of Physiology & Pathophysiology,

Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba.

Winnipeg, MB R3E 0J9, Canada

Correspondence:

Etienne Leygue, E-mail: Etienne.Leygue@umanitoba.ca, 5029 - 675 McDermot Ave., Winnipeg, MB R3E 0V9, Canada. Phone: (204) 787-2785, Fax: (204) 787-2190

Jiuyong Xie, E-mail: xiej@umanitoba.ca, 440 BMSB, 745 Bannatyne Ave., Winnipeg, MB R3E 0J9, Canada. Phone: 204 9757774, FAX: 204 789 3934.

Abstract:

We discuss the term “long non-coding RNA” in light of the discovery of peptides encoded by some of them, interesting questions arising from the discovery and propose to use “long non-protein-coding RNA” instead.

Keywords: lncRNA, peptides, protein

Long 'non-coding' RNAs (lncRNAs) were originally defined as RNA longer than 200nt but exempt from any open reading frame potentially leading to the production of a protein (Nagano and Fraser 2009; Rinn and Chang 2012). Thousands of these lncRNA genes have now been identified (Ulitsky and Bartel 2013).

Some confusion has however been arising surrounding the adequacy of using the term 'non-coding' to qualify some of these lncRNAs. Indeed, some translation products, even though shorter than most proteins, might exist and be biologically important.

Functional short peptides are not something new. They include peptidic hormones such as insulin (51aa), adrenocorticotropin (ACTH, 39aa) and melanocyte-stimulating hormones(12-22aa); even though the latter two are processed from longer precursors. Short peptides have been proven not only to fulfill potent and important physiological functions (Koeppen et al. 2010), but also to act as lethal toxins (Gray et al. 1988; Becerril et al. 1996; Lewis and Garcia 2003).

It is therefore not surprising in this regard that recently identified short peptides encoded by two 'lncRNAs' have functional relevance in human and mouse cells(Anderson et al. 2015; Nelson et al. 2016).

In their articles, the authors demonstrated the existence of two peptides, myoregulin (MLN) and dwarf open reading frame (DWORF), 46aa and 34aa long, which are specifically expressed in skeletal muscle and heart/soleus tissues, respectively. Using the CRISPR/Cas9 system to disrupt the ORF (KO), they found that loss of DWORF peptide expression, demonstrated by an absence of specific signal

46 detectable by Western blot analysis of muscle proteins from homozygous mice, was paralleled by a
47 reduction SERCA effect on the calcium transients.

48 These observations raise several important questions regarding our current definition of lncRNA
49 or even ncRNA in general.

50 Are the functions currently attributed to known characterized lncRNAs solely depending on the
51 RNA molecule? In other words, could some of the peptides potentially encoded by these transcripts
52 contribute to the function observed, alone or in combination with the lncRNA considered?

53 The existence of genes leading to the production of a functional RNA as well as a protein have
54 been documented in multiple organisms(Ulveling et al. 2011). In drosophila for instance, the *OSKAR*
55 gene produces RNA sequences that do not require translation to regulate oocyte development(Jenny
56 et al. 2006), as well as a protein involved in the spatial distribution of its own RNA (Rongo et al. 1995;
57 Ryu and Macdonald 2015). In mammals, the Steroid Receptor RNA Activator gene (*SRA1*) generates a
58 functional non-coding RNA (SRA)(Lanz et al. 1999) involved in the regulation of multiple transcription
59 factors as well as a protein (SRAP), recently found to control the motility of cancer cell(Chooniedass-
60 Kothari et al. 2004; Leygue 2007; McKay et al. 2014; Yan et al. 2015).

61 The bi-faceted aspect of such systems, i.e the co-existence of a functional RNA and a protein,
62 obviously complicates the overall interpretation of experiments performed. It emphasizes the need for
63 designing experimental procedures that could clearly delineate the function of either protagonist.
64 When considering the RNA aspect of the system, it is therefore important to demonstrate function in
65 the absence of any translation product. For example, such approaches as treating cells with
66 cycloheximide or introducing mutations altering putative start codons or introducing multiple stop
67 codons have been successfully used to establish the functional role of SRA(Lanz et al. 1999).

Interestingly, experiments involving silent mutations not altering the peptide sequence can be interpreted to validate either the RNA function(Lanz et al. 2002) or the protein function(Chooniedass-Kothari et al. 2010; Yan et al. 2015), depending on the experimental context. More importantly, evidence from functional rescue with either RNA or peptide in cells deleted of a lcnRNA gene will greatly help clarify their respective functions.

Are there more lncRNAs in the genome that encode peptides? Alba and colleagues have shown in 2014 that 28.6% of lncRNAs in yeast and 81.9% in humans are associated with ribosomes (Ruiz-Orera et al. 2014). If even a portion of these 'lncRNAs' are destined for translation, this could dramatically increase the wealth of coding potential of genomes/transcriptomes. Interestingly, Ulveling *et al.* have considered the issue through an opposite angle, mostly by trying to identify open reading frames which would be interrupted by retention of short introns(Ulveling et al. 2011). They hypothesized that such transcripts might have functions remaining to be determined.

How many peptides per lncRNA could be potentially generated? Considering that peptides potentially able to function as hormones or even toxins(Possani et al. 2000) can be short (30aa), it is reasonable to suggest that more than one peptide could be encoded by some 'lncRNAs', further adding to the complexity of the whole population of regulatory molecules. These peptides could be produced by either separate or single translation (for instance, followed by cleavage as in POMC(Koeppen et al. 2010)).

Corollar to the existence of the peptides encoded by transcripts most still referred to as lncRNAs, lay questions relating to their putative role in evolution and most importantly their implication in diseases. It is indeed likely that mutations affecting lncRNAs, affecting RNA structures hence their function or their peptidic outcome could alter phenotypes and cause diseases. As such, mutations in

the *SRA1* gene, encompassing regions evolutionarily well conserved of the transcript(Novikova et al. 2012), as well as SRAP(Cooper et al. 2011), have recently been pinpointed in families affected by idiopathic hypogonadotrophic hypogonadism (IHH)(Kotan et al. 2016).

It is urgent to investigate throroughly the existence of peptides potentially encoded by lncRNAs. As the size/sequence of these molecules might preclude the use of classical tool, their detection in blood or other fluid as well as in tissues should remain a priority for proteomic researchers. One might indeed foresee that these peptides have the potential to be used as markers of diseases, therapeutic targets or even drugs.

Considering the current nomenclature, we suggest herein that 'lncRNA' is no longer adequate to tag these transcripts that can encode and be translated for peptides. The term 'lncRNA' should refer to 'long non-protein-coding RNA' rather than 'non-coding RNA'. In this context protein would mean polypeptides with a molecular weight of more than 10kD according to the IUPAC guide(McNaught et al. 2000). This should cover most, if not all, lncRNAs at present.

105 **References**:

106 Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P.,
107 Shelton, J.M., Liou, J., Bassel-Duby, R., and Olson, E.N. 2015. A micropeptide encoded by a putative long
108 noncoding RNA regulates muscle performance. *Cell* **160**(4): 595-606.
109 Becerril, B., Corona, M., Coronas, F.I., Zamudio, F., Calderon-Aranda, E.S., Fletcher, P.L., Jr., Martin, B.M., and
110 Possani, L.D. 1996. Toxic peptides and genes encoding toxin gamma of the Brazilian scorpions Tityus
111 bahiensis and Tityus stigmurus. *The Biochemical journal* **313 (Pt 3)**: 753-760.
112 Chooniedass-Kothari, S., Emberley, E., Hamedani, M.K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M.,
113 Watson, P.H., and Leygue, E. 2004. The steroid receptor RNA activator is the first functional RNA
114 encoding a protein. *FEBS letters* **566**(1-3): 43-47.
115 Chooniedass-Kothari, S., Vincett, D., Yan, Y., Cooper, C., Hamedani, M.K., Myal, Y., and Leygue, E. 2010. The
116 protein encoded by the functional steroid receptor RNA activator is a new modulator of ER alpha
117 transcriptional activity. *FEBS letters* **584**(6): 1174-1180.
118 Cooper, C., Vincett, D., Yan, Y., Hamedani, M.K., Myal, Y., and Leygue, E. 2011. Steroid Receptor RNA Activator
119 bi-faceted genetic system: Heads or Tails? *Biochimie* **93**(11): 1973-1980.
120 Gray, W.R., Olivera, B.M., and Cruz, L.J. 1988. Peptide toxins from venomous Conus snails. *Annual review of*
121 *biochemistry* **57**: 665-700.
122 Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdelyi, M., and Ephrussi, A. 2006.
123 A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development* **133**(15): 2827-
124 2833.
125 Koeppen, B.M., Stanton, B.A., and Berne, R.M. 2010. *Berne & Levy physiology*. Mosby/Elsevier, Philadelphia, PA.
126 Kotan, L.D., Cooper, C., Darcan, S., Carr, I.M., Ozen, S., Yan, Y., Hamedani, M.K., Gurbuz, F., Mengen, E., Turan, I.,
127 Ulubay, A., Akkus, G., Yuksel, B., Topaloglu, A.K., and Leygue, E. 2016. Idiopathic Hypogonadotropic
128 Hypogonadism Caused by Inactivating Mutations in SRA1. *Journal of clinical research in pediatric*
129 *endocrinology*.
130 Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. 1999. A
131 steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**(1): 17-
132 27.
133 Lanz, R.B., Razani, B., Goldberg, A.D., and O'Malley, B.W. 2002. Distinct RNA motifs are important for
134 coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proceedings of the*
135 *National Academy of Sciences of the United States of America* **99**(25): 16081-16086.
136 Lewis, R.J. and Garcia, M.L. 2003. Therapeutic potential of venom peptides. *Nature reviews Drug discovery* **2**(10):
137 790-802.
138 Leygue, E. 2007. Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected
139 relevance to breast cancer. *Nuclear receptor signaling* **5**: e006.
140 McKay, D.B., Xi, L., Barthel, K.K., and Cech, T.R. 2014. Structure and function of steroid receptor RNA activator
141 protein, the proposed partner of SRA noncoding RNA. *Journal of molecular biology* **426**(8): 1766-1785.
142 McNaught, A.D., Wilkinson, A., International Union of Pure and Applied Chemistry., and Royal Society of
143 Chemistry (Great Britain). 2000. IUPAC compendium of chemical terminology. In. Royal Society of
144 Chemistry,, Cambridge, England.
145 Nagano, T. and Fraser, P. 2009. Emerging similarities in epigenetic gene silencing by long noncoding RNAs.
146 *Mammalian genome : official journal of the International Mammalian Genome Society* **20**(9-10): 557-562.
147 Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R.,
148 Chen, X., Kavalali, E.T., Cannon, S.C., Houser, S.R., Bassel-Duby, R., and Olson, E.N. 2016. A peptide
149 encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*
150 **351**(6270): 271-275.
151 Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. 2012. Structural architecture of the human long non-coding
152 RNA, steroid receptor RNA activator. *Nucleic acids research* **40**(11): 5034-5051.

Possani, L.D., Merino, E., Corona, M., Bolivar, F., and Becerril, B. 2000. Peptides and genes coding for scorpion toxins that affect ion-channels. *Biochimie* **82**(9-10): 861-868.

Rinn, J.L. and Chang, H.Y. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**: 145-166.

Rongo, C., Gavis, E.R., and Lehmann, R. 1995. Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Development* **121**(9): 2737-2746.

Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. 2014. Long non-coding RNAs as a source of new peptides. *eLife* **3**: e03523.

Ryu, Y.H. and Macdonald, P.M. 2015. RNA sequences required for the noncoding function of oskar RNA also mediate regulation of Oskar protein expression by Bicoid Stability Factor. *Developmental Biology* **407**(2): 211-223.

Ulitsky, I. and Bartel, D.P. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**(1): 26-46.

Ulveling, D., Francastel, C., and Hube, F. 2011. When one is better than two: RNA with dual functions. *Biochimie* **93**(4): 633-644.

Yan, Y., Cooper, C., Hamedani, M.K., Guppy, B., Xu, W., Tsuyuki, D., Zhang, C., Nugent, Z., Blanchard, A., Davie, J.R., McManus, K., Murphy, L.C., Myal, Y., and Leygue, E. 2015. The steroid receptor RNA activator protein (SRAP) controls cancer cell migration/motility. *FEBS letters* **589**(24 Pt B): 4010-4018.

Last version on June 6, 2016