UTR and non-coding RNA: reconnecting terms to function

Adeilton Brandão

Instituto Oswaldo Cruz - Fiocruz, Rio de Janeiro, Brasil

Corresponding Author:

Adeilton Brandão

email address: abrandao@fiocruz.br

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3102v1 | CC BY 4.0 Open Access | rec: 21 Jul 2017, publ: 21 Jul 2017

ABSTRACT

Scientific terms should be as accurate and meaningful as possible for both researchers and the general science readership. Currently, some scientific terms do not properly describe the activity or function to which they are associated to, being frequently characterized by negative reference to a prior feature or finding. UTR (Untranslated Region) and non-coding RNA fall within this class. In this article, I argue for a revision of these terms to account for the growing lines of evidence about their known function and activity in the cell.

KEYWORDS: UTR; untranslated region; non-coding RNA; translation; transcription; scientific names; biological coding

INTRODUCTION

Scientific names should be based on words that could offer an accurate and quick grasp of their intended meanings. If the initial designation or abbreviation for a scientific finding is adequately formulated, it can accompany future advancements in such way that the accumulation of new evidence to its knowledge base would not require the name to be constantly corrected. An example of such a designation is the term “messenger RNA” (mRNA). Its description in the 1960's (Gros et al., 1961) resulted in the choice of a name that, even after a torrent of facts about its structure, composition and activity, remains quite appropriate nowadays.

The same cannot be said about the 5' and 3' Untranslated Region (5' and 3' UTR). The first reference to “untranslated” segments appeared in 1970 from research on the R17 bacteriophage (Adams & Cory, 1970). They were more precisely described from the sequencing of the rabbit beta globin cDNA in 1977 (Efstratiadis et al., 1977), which demonstrated that the 5' and 3' ends of the mRNA did not match the protein amino acid sequence, leading to the conclusion that they do not contribute to the primary sequence of the translated polypeptide. The term Untranslated Region remained unquestioned until the discovery in 1991 of the small open reading frames observed in the 5' UTR of some genes that might also be translated (Abastado et al., 1991). From this moment on, the term 'Untranslated Region' no longer accurately described all sequences of an mRNA that precede the main start codon. Though not as frequent as the 5' end, the 3' end of some mRNAs may also harbor small open reading frames

(Mackowiak et al., 2015). Despite new evidence contradicting the contemporary wisdom, the 5' and 3' ends of an mRNA continued to be called UTRs, a term that could incorrectly associate them to a behavior they might not display.

Not all 5' UTRs have peptide encoding capability: only those displaying a small open reading frame (ORF), also known as upstream ORF (uORF), might be translated. Furthermore, a uORF does not need to be translated to display an activity. Its sole presence in the 5' UTR might be sufficient to influence the translation of the main ORF in the respective mRNA (Mueller & Hinnebusch, 1986; Gaba et al., 2001). Researchers have also demonstrated that the 3' UTRs are involved in the control of gene expression, translation regulation, mRNA stability, localization, turn over, micro RNA binding, protein-protein interaction (Mayr, 2016; Lai, 2002).

NEW FUNCTIONS, MORE CODES, NEW TERMS.

With so many activities should the 5' and 3' ends of an mRNA be still called the "Untranslated Region"? The answer to this question should include an alternative view: beyond the translation potential, the mRNA ends have other coding capacities, *i. e.*, they have implicit codes that signalize binding regions for proteins and factors involved in a myriad of interactions in the cell. Referring to 5' and 3' ends of mRNA as simply "Untranslated Region" gives no hint about their multiple functions. Indeed, it is misleading, as the words "Untranslated Region" implies that the sequences in these segments have no peptide coding capacity and obfuscates any other role in mRNA metabolism.

The current information about the importance of the mRNA ends for cell metabolism suggests that we need a more appropriate designation for these segments. Considering that the 5' and 3' ends of mRNA do not appear in the final protein, yet participate in some way in the translation process, and that this participation depends on the implicit codes contained within the nucleotide composition, we should think of new words that better reflect this phenomenon. We may describe them as 5' or 3' "**Hyper Coding Segments**" (5' HCS or 3' HCS) to contrast them to the often more prominent mRNA segment that codes for a protein. Under this designation, we would be stating that these segments are potentially capable of a larger set of actions dependent on nucleotide composition either at the primary or secondary level, which might result in:

- translated peptides, (presence of a uORF).
- secondary structures that can influence translation or mRNA dynamics.

- recognition/binding segments (AU binding proteins; riboswitches in 5' end, miRNA in 3' end).
- protein-protein interaction mediation.
- alternative splicing (cis or trans splicing).
- alternative polyadenylation.

The denomination suggested above does not explicitly state an activity or its absence/presence, but instead provides an ample definition: an mRNA component separated from the main protein encoding segment whose sequences provide variable coding content with the potential to influence several cellular activities.

Another case is the designation for RNAs that do not fall within current categories, *i. e.*, mRNA, tRNA, rRNA, miRNA, siRNA, snRNA snoRNA, SL RNA, gRNA. In the absence of a better functional definition, RNAs, which vary from tens to several hundreds of bases in length, have been collectively designated as the non-coding RNA (ncRNA) (Eddy, 2002). This terminology does not comprehend the generalized role RNA plays in cellular processes, and denies the importance these molecules have in the widening of the biological code concept. Indeed, it maintains the attachment to the code concept from the first years of molecular biology, which was restricted to an information flow from nucleic acid to peptides, *i. e.* "translating a DNA molecule into a polypeptide having an RNA molecule as intermediary".

From this point of view, an RNA that cannot be inserted into one of the categories mentioned above apparently implies that it does not code for or convey any information. A decade ago, Gingeras has suggested that this type of RNA should be referred as TUF, "*Transcripts of Unknown Function*" (Gingeras, 2007). Though this designation precisely describes what this molecule is ("a transcript"), it does not explicitly make reference to a capacity for coding biological information.

Recent research results signalize some change about the coding content of the “non-coding” RNA (Ruiz-Orera et al., 2017; Atkinson et al., 2017). We might expect that there is information conveyed by these molecules, and there might be an implicit code in their composition. Thus, we should not use the terminology of 'non-coding RNA" or "transcript of unknown function". While we are unable to fully predict what all this code signifies to cell function, if we assume that there are multiple information layers within the sequences of these RNAs, then a more appropriate reference to them could be "***meta code"*** RNA. This designation would imply that there are other codes

beyond the conventional polypeptide view of biological encoding.

A review of these terms should start by moving from negative association to an updating of the coding concept, which has an important implication for the biology research in general: any DNA segment in a given genome is a carrier of coded information, and as consequence, every genome is interlaced with codes.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGMENTS

I thank David Provance Jr. and Patricia Cuervo for their helpful suggestions and comments.

REFERENCES

Abastado JP, Miller PF, Jackson BM, Hinnebusch AG. 1991. Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Molecular and Cellular Biology* 11:486-496.

Adams JM, Cory S. 1970. Untranslated nucleotide sequence at the 5'-end of R17 bacteriophage RNA. *Nature* 227:570-574.

Atkinson SR, Marguerat S, Bitton DA, Rodríguez-López M, Rallis C, Lemay JF, Cotobal C, Malecki M, Mata J, Bachand F, Bahler J. 2017. Long non-coding RNA repertoire and regulation by nuclear exosome, cytoplasmic exonuclease and RNAi in fission yeast. *bioRxiv* 158477; doi: https://doi.org/10.1101/158477.

Eddy SR. 2002. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* 2:919-929.

Efstratiadis A, Kafatos FC, Maniatis T. 1977. The primary structure of rabbit beta-globin mRNA as determined from cloned DNA. *Cell* 10:571-585.

Gaba A, Wang Z, Krishnamoorthy T, Hinnebusch AG, Sachs MS. 2001. Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *The EMBO Journal* 20:6453-6463.

Gingeras TR. 2007. Origin of phenotypes: genes and transcripts. *Genome Research* 17:682-690.

Gros F, Gilbert W, Hiatt H, Attardi G, Spahr P, Watson J. 1961. Molecular and biological characterization of messenger RNA. *Cold Spring Harbor Symposia on Quantitative Biology* 26:111-132.

Lai EC. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics* 30:363-364.

Mackowiak SD, Zauber H, Bielow C, Thiel D, Kutz K, Calviello L, Mastrobuoni G, Rajewsky N, Kempa S, Selbach M, Obermayer B. 2015. Extensive identification and analysis of conserved small ORFs in animals. *Genome Biology* 16:179.

Mayr C. 2016. Evolution and Biological Roles of Alternative 3′UTRs. *Trends in Cell Biology* 26:227–237.

Mueller PP, Hinnebusch AG. 1986. Multiple upstream AUG codons mediate translational control of GCN4. *Cell* 45:201-207.

Ruiz-Orera j, Verdaguer-Grau P, Villanueva-Cañas JL, Messeguer X, Albà MM. 2017. Evidence for functional and non-functional classes of peptides translated from long non-coding RNAs. *bioRxiv* 064915; doi: https://doi.org/10.1101/064915