

Genetic variation may play a crucial role in non-coding RNA biogenesis

Jeyalakshmi Kandhavelu¹, Meenakshisundaram Kandhavelu^{2*}

¹School of Veterinary Medicine, University of Camerino, Camerino, Italy

²Department of Signal Processing, Tampere University of Technology, Tampere, Finland

Email: meenakshisundaram.kandhavelu@tut.fi

Received 29 June 2012; revised 31 July 2012; accepted 24 August 2012

ABSTRACT

Transcription, post-transcriptional modification, translation, post-translational modification, DNA replication, and signaling interaction of intra- and extracellular components are the relevant mechanisms in gene regulation. Transcription is one of the most important mechanisms in the control of gene expression. Further, post-transcriptional modifications play a crucial role after transcription which determine whether the transcribed gene is coding or non-coding RNA (ncRNAs). Genome-wide analysis of RNAs provides information about the coding RNAs, whereas the status of ncRNAs are still at large and must be discussed in detail as variations in the ncRNAs can lead to different phenotypes. In this short article, we discuss the role of genetic variation in ncRNA genes and how this variation may play a crucial role in ncRNA biogenesis that eventually leads to phenotypic variation and thus speciation.

Keywords: Genome Wide Analysis; Bioinformatics; Genetic Variation; Non-Coding RNA Biogenesis; RNA Structure; Speciation

1. DARK MATTERS IN GENE EXPRESSION

Our understanding of the principle mechanisms that orchestrate the central dogma of life is limited due to the lack of appropriate physical approaches [1,2], *i.e.* accurate and advanced experimental facilities to understand the genetic variations, the limitations in the bioinformatics tools which predict DNA/RNA secondary structure, interactions, control elements, among others. Due to these limitations, a detailed understanding of the exact mechanisms underlying the processes of transcription and translation are still obscure. Transcription is a complex and tightly regulated mechanism in gene expression where

the genes can be turned on or off based on internal or external signals [3]. The post transcriptional modifications of the transcripts designates them as coding (protein-coding) or non-coding RNAs (ncRNAs) [4-6]. Extensive research on RNA proves their significant role in gene expression [7,8]. Recently, a pilot study reported the properties of loss-of-function (LoF) variants of human protein-coding genes [9], which show that each human individual carries ~100 LoF variants, with ~20 genes completely lost. These loci can be described as “dark matter” because intronic sequences hide crucial information in their sequences which affect gene expression. There is evidence that variations in the exonic or the intronic regions of DNA can affect the structure and functions of their target mRNA or ncRNAs and vice versa, and thus the cellular signaling pathways of the cells that control, for example, the birth and death of cells. It is thus important to discuss the status of genetic variations not only in coding RNA but also in the ncRNAs. While ncRNAs are important in the control of gene expression, it is unclear how these ncRNAs originate. Below, we discuss the role of genetic variation in creating variation in ncRNAs during biogenesis and how it affects the structure of ncRNAs.

2. ROLE OF GENETIC VARIATION IN NCRNA BIOGENESIS

Each individual’s genome differs in its expression patterns from other members of the species, as well as between species [10]. Using a Markov model, one recent report suggests that humans and chimpanzees speciated 4.1 million years ago. Total human genome analysis reported that only <2% of the total genome sequence has protein-coding capacity [11]. This implies that ~98% of the genome might have both coding and ncRNA producing genes, which might show expression variation in any given environment. These variations would likely impart changes in two major components: 1) it could affect the secondary structure of the encoded protein (as a result, LoF variant proteins could be generated as shown by

*Corresponding author.

MacArthur *et al.*), and 2) it could mis-target the ncRNAs by relaxing the base pairing consistency [12]. Such mis-pairing might be more likely to affect the function of particular RNAs, and could also disturb cellular signaling pathways controlling gene expression, since ncRNAs regulate the expression of many genes [13]. It would be more interesting if genome sequence data also discussed the LoF variants of intronic sequences. This would help us understand whether the predicted LoF variants were generated by intronic variation. Recent evidence also suggests that certain sequences have the potential to generate ncRNAs with different functions (possibly through non-canonical ncRNA biogenesis pathways [14]), which are important for cellular functions. This includes Sno-microRNA [15] and pi-sno-microRNA (x-ncRNA) [16,17], which are evolutionarily conserved. This leads to the intriguing question: how does a particular gene sequence produce structurally and functionally different ncRNAs (see **Figure 1**). The answer for this phenomenal question is currently unknown and the approaches to address this question are being recognized as an active area of investigation. Whether or not the single-nucleotide polymorphism (SNP) variants of ncRNA could play crucial roles also remains to be clarified.

The primary ncRNA sequences are transcribed from DNA sequences [2]. It is not known how deletion, addition, repeats and other LoF variation in the DNA sequence of ncRNAs affect the structure of the primary transcript [18]. Genetic variation may play a critical role,

not only in the LoF of mRNAs but also in ncRNAs. Although experimental evidence suggests that several proteins (e.g. Dicer, RISC) are involved in the production of a particular ncRNA from precursor-ncRNA [19,20], this evidence could not elucidate the secondary or tertiary structure of the primary transcripts. This is because of the stochasticity in the formation of the tertiary RNA structure which includes energetic formations (Gibbs theory of free energy function) of tertiary structure, chelation of divalent cations/multicharged ligand interaction in RNA sequences which are poorly understood [21,22]. Data analysis of LoF variants of coding genes confirm the presence of changes in the sequence of RNA and hence changes in the structure and function of the RNA. This phenomenon hints that ncRNA might be more susceptible to structural and functional changes due to changes in sequence [23] if the aforementioned RNA folding theories are true. It is thus also important to address the status of an ncRNA variant in the same environmental condition in any human genome data, since it might give a path to predict the biogenesis pathway of ncRNA families. For example, in the case of Cystic fibrosis, LoF variants of non-coding genes have deleterious effects [24]. Since most of the human genome accumulates and transmits genetic mutations over time, understanding the detailed status of the noncoding sequence mutations along with the genetic variation of coding sequences might help us overcome deleterious effects [25] and might allow us to see ongoing genetic evolution in this era.

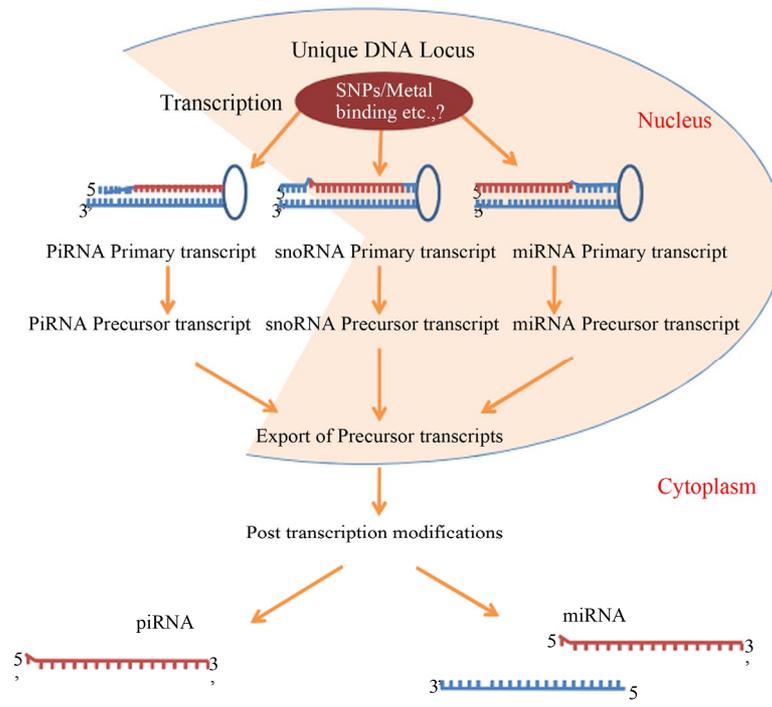


Figure 1. SNPs/metal induced RNA structural variations produce different types of non-coding RNAs from the same locus.

3. THEORIES OF GENETIC VARIATION, LOF AND SPECIATION

Theories on genetic variations have been discussed extensively for several decades. Recently, a report reviewed several proposed the theories of genetic variations [26] as: 1) “Less is more” hypothesis—advantageous effects of LoF variants; 2) “Less is less” hypothesis—deleterious effects of LoF variants; and 3) “less is nothing”—tolerated LoF variants situation. These theories describe the LoF of particular characteristics based on the comparison of populations or by tracking of an individual trait. The adaptation of a sub-population leads to the evolution of new species. This means that adopted characteristics are transferred to offspring which undergo natural selection, where the favorable traits are preserved allowing the sub-population to survive, which may then be subject to speciation. Further, Charles Darwin proposed that “each slight variation, if useful, is preserved” which determines how a sub-population adapts and evolves to form a new species in the process of natural selection [27].

A systematic survey of LoF variant genes using clinical sequencing data provides the status of individuals affected by a particular disease and infers that the selected individuals experience the deleterious effect by LoF variants in that particular environment [28]. While discussing the deleterious effects of genetic variation, it is important to extract more information. For example, it is not known whether these effects may become advantageous to the offspring, if transmitted. Further, the position of the mutation in the intronic gene sequence might play a significant role certain diseases. Most genome-wide data lacks single nucleotide polymorphisms (SNPs) of introns. Several reports provide evidence that SNPs can affect the structure and function of ncRNAs, which can cause diseases such as cancer [29]. Providing this data will add much more value to understand how the “less is less hypothesis” plays a role in the selection of cells in a population by coordinating gene expression patterns via coding and ncRNAs.

The LoF variants in ncRNAs may play a significant role in the biogenesis of a particular type of ncRNA and hence lead to the production of multiple types of ncRNA from the same sequence. LoF variants in coding and ncRNA may thus control gene expression of cells in a given environment in a combinatorial manner for its survival. In human genome data, understanding the LoF variant function in ncRNA might therefore help us explore the real status of the gene expression system in a particular environment. This will allow us to study how the less is less (the role of deleterious effects in speciation) phenomenon occurs in the natural system.

REFERENCES

- [1] Crick, F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561-563. [doi:10.1038/227561a0](https://doi.org/10.1038/227561a0)
- [2] Mattick, J.S. (2003) Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930-939. [doi:10.1002/bies.10332](https://doi.org/10.1002/bies.10332)
- [3] Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research*, **11**, 1475-1489. [doi:10.1093/nar/11.5.1475](https://doi.org/10.1093/nar/11.5.1475)
- [4] Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, **2**, 919-929. [doi:10.1038/35103511](https://doi.org/10.1038/35103511)
- [5] Wilusz, J.E., Sunwoo, H. and Spector, D.L. (2009) Long noncoding RNAs: Functional surprises from the RNA world. *Genes & Development*, **23**, 1494-1504. [doi:10.1101/gad.1800909](https://doi.org/10.1101/gad.1800909)
- [6] Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*, **10**, 155-159. [doi:10.1038/nrg2521](https://doi.org/10.1038/nrg2521)
- [7] He, L. and Hannon, G.J. (2009) MicroRNAs: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, **5**, 522-531. [doi:10.1038/nrg1379](https://doi.org/10.1038/nrg1379)
- [8] McKnight, S.L. and Kingsbury, R. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. *Science*, **217**, 316-324. [doi:10.1126/science.6283634](https://doi.org/10.1126/science.6283634)
- [9] MacArthur, D.G., Balasubramanian, S., Frankish, A. *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, **335**, 823-828. [doi:10.1126/science.1215040](https://doi.org/10.1126/science.1215040)
- [10] Mayr, E. (1963) Animal species and evolution. Belknap Press of Harvard University Press, Cambridge.
- [11] Hobolth, A., Christensen, O.F., Mailund, T. *et al.* (2007) Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLOS Genetics*, **3**, e7. [doi:10.1371/journal.pgen.0030007](https://doi.org/10.1371/journal.pgen.0030007)
- [12] Ritz, J., Martin, J.S. and Laederach, A. (2012) Evaluating our ability to predict the structural disruption of RNA by SNPs. *BMC Genomics*, **13**, S6. [doi:10.1186/1471-2164-13-S4-S6](https://doi.org/10.1186/1471-2164-13-S4-S6)
- [13] Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Human Molecular Genetics*, **15**, 17-29. [doi:10.1093/hmg/ddl046](https://doi.org/10.1093/hmg/ddl046)
- [14] Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology*, **10**, 126-139. [doi:10.1038/nrm2632](https://doi.org/10.1038/nrm2632)
- [15] Scott, M.S., Avolio, F., Ono, M., *et al.* (2009) Human miRNA precursors with box H/ACA snoRNA features. *PLOS Computational Biology*, **5**, e1000507.
- [16] Kandhavelu, M., Lammi, C., Buccioni, M., *et al.* (2009) Existence of snoRNA, microRNA, piRNA characteristics in a novel non-coding RNA: X-ncRNA and its biological implication in *Homo sapiens*. *Journal of Bioinformatics*

and Sequence Analysis, **1**, 31-40.

- [17] Kandhavelu, M. and Kandhavelu, J. (2012) Pre-piRNA biogenesis mimics the pathway of miRNA. *Biochemical Systematics and Ecology*, **43**, 200-204. [doi:10.1016/j.bse.2012.03.012](https://doi.org/10.1016/j.bse.2012.03.012)
- [18] Eddy, S.R. (2002) Computational genomics of noncoding RNA genes. *Cell*, **109**, 137-140. [doi:10.1016/S0092-8674\(02\)00727-4](https://doi.org/10.1016/S0092-8674(02)00727-4)
- [19] Hutvagner, G., McLachlan, J., Pasquinelli, A.E., *et al.* (2001) A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834-838. [doi:10.1126/science.1062961](https://doi.org/10.1126/science.1062961)
- [20] Gregory, R.I., Chendrimada, T.P., Cooch, N., *et al.* (2005) Human RISC couples microRNA biogenesis and post-transcriptional gene silencing. *Cell*, **123**, 631-640. [doi:10.1016/j.cell.2005.10.022](https://doi.org/10.1016/j.cell.2005.10.022)
- [21] Draper, D.E. (2004) A guide to ions and RNA structure. *Rna—A Publication of the Rna Society*, **10**, 335-343.
- [22] Draper, D.E. (2008) RNA folding: Thermodynamic and molecular descriptions of the roles of ions. *Biophysical Journal*, **95**, 5489-5495. [doi:10.1529/biophysj.108.131813](https://doi.org/10.1529/biophysj.108.131813)
- [23] Mathews, D.H., Sabina, J., Zuker, M., *et al.* (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, **288**, 911-940. [doi:10.1006/jmbi.1999.2700](https://doi.org/10.1006/jmbi.1999.2700)
- [24] Chillon, M., Dork, T., Casals, T., *et al.* (1995) A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA-->G, produces a new exon: High frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *The American Journal of Human Genetics*, **56**, 623-629.
- [25] Montgomery, S.B. and Dermitzakis, E.T. (2011) From expression QTLs to personalized transcriptomics. *Nature Reviews Genetics*, **12**, 277-282. [doi:10.1038/nrg2969](https://doi.org/10.1038/nrg2969)
- [26] Quintana-Murci, L. (2012) Genetics. Gene losses in the human genome. *Science*, **335**, 806-807. [doi:10.1126/science.1219299](https://doi.org/10.1126/science.1219299)
- [27] Darwin, C. (1859) On the origin of species by means of natural selection. J. Murray, London.
- [28] Cairns, J. (1975) Mutation selection and the natural history of cancer. *Nature*, **255**, 197-200. [doi:10.1038/255197a0](https://doi.org/10.1038/255197a0)
- [29] Garzon, R., Calin, G.A. and Croce, C.M. (2009) MicroRNAs in cancer. *Annual Review of Medicine*, **60**, 167-179. [doi:10.1146/annurev.med.59.053006.104707](https://doi.org/10.1146/annurev.med.59.053006.104707)