The so-called non-coding RNAs are not really non-coding after all

The three major species of cellular RNA molecules – mRNA, tRNA and rRNA – were identified during the golden era of molecular biology in the 1960s and 70s as important mediators in deciphering genetic information in the DNA. Among them, mRNA is the only 'coding' species that carries sequential genetic codes that are translated to the amino acid sequence of proteins. On the other hand, tRNA and rRNA are classified as 'non-coding' RNA (ncRNA) that assists in mRNA translation. A fourth class of RNA called 'spliceosomal RNA' was discovered later, which is also non-coding in nature and is involved in the processing of mRNA. Though tRNA, rRNA and spliceosomal RNA do not directly code for proteins, they are indirectly involved in relaying information from DNA to proteins. A quantum jump in RNA research was made at the turn of this millennium with the discovery of a new class of ncRNA molecules that are regulatory in nature and can control gene expression. Over the last decade and a half, research on these RNAs has grown exponentially. With the genome and transcriptome sequence information available from increasing number of organisms, we now know that only a small fraction of the genome (<2% in humans) codes for proteins, even though the vast majority of the remaining genome is also transcriptionally active. Such transcriptional activity of the intergenic region, which was earlier thought to be non-coding 'junk DNA', results in the generation of a myriad of ncRNA of diverse sequence and length, the functions of which have been debated, or even doubted, by many for years. Based on their size, these RNAs are broadly divided into long ncRNA (lncRNA) and short ncRNA that includes microRNA (miRNA). In today's parlance, ncRNA primarily refers to these classes of RNA.

Despite their nomenclature, recent work shows that these so-called non-coding RNA molecules do possess coding capacity after all. Even though they are not capable of encoding long peptides that fold into functional proteins such as enzymes or transcription factors, they code for short peptides termed as 'micropeptides' that play an important regulatory role in developmental and cellular processes. The first example came in 2010, when it was shown that a putative lncRNA named 'polished rice' in the fruit fly encodes four homologous micropeptides of 11–32 residues length that regulate the formation of trichomes, the hair-like epidermal cells projecting off the skin surface[1]. They do this by initiating an amino-terminal truncation of a known transcription factor called shaven baby, via ubiquitin-mediated proteasomal pathway, thereby converting it from a repressor to an activator protein. More recently, an lncRNA-encoded micropeptide of 46 residue length, called myoregulin, has been identified in mammals that regulates twitching of muscles[2]. Removal of this lncRNA gene from the genome resulted in improved exercise performance in transgenic mice.

The real surprise on the coding potential of the ncRNA world came a few months ago from a study on plant miRNAs. As mentioned above, miRNAs are short regulatory RNA molecules that inhibit gene expression either by directly binding the target mRNA molecules and degrading them, or by blocking their translation into proteins. miRNAs are initially transcribed as long transcripts called primary miRNAs (pri-miRNAs), which are eventually processed into 21–22 nucleotide long mature miRNAs that, in combination with certain proteins, are capable of inhibiting their target transcripts. Until now, pri-miRNAs were thought to be only serving as the precursors for the mature miRNAs. A recent paper in *Nature*[3] has dramatically changed this notion. It shows that the primary transcripts of miRNAs in *Arabidopsis* and alfalfa, apart from providing mature miRNAs, also encode micropeptides (aptly named miPEPs) that play a key regulatory role in organ development.

The discovery of the coding capability of these RNA molecules that were earlier classified as non-coding with reference to proteins opens a new area of research involving micro peptides[4,5] and their role in cellular regulation. The therapeutic potential of exogenous addition of such regulatory micropeptides needs to be examined.

1. Kondo, T. *et al.*, *Science*, 2010, **329**, 336–339.
2. Anderson, D. M. *et al.*, *Cell*, 2015, **160**, 595–606.
3. Lauressergues, D., Couzigou, J. M., Clemente, H. S., Martinez, Y., Dunand, C., Bécard, G. and Combier, J. P., *Nature*, 2015, **520**, 90–93.
4. Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A. and Couso, J. P., *PLoS Biol.*, 2007, **5**, e106.
5. Hashimoto, Y., Kondo, T. and Kageyama, Y., *Dev. Growth Differ. (Suppl. 1)*, 2008, **50**, S269–S276.

Shree Chaitranjali Yadla and **Utpal Nath*,** Department of Microbiology and Cell Biology, Indian Institute of Science, Bengaluru 560 012, India.
*e-mail: utpal@mcbl.iisc.ernet.in