Some long non-coding RNAs are conventional after all

April 4 2014

Not so long ago researchers thought that RNAs came in two types: coding RNAs that make proteins and non-coding RNAs that have structural roles. Then came the discovery of small RNAs that opened up whole new areas of research. Now researchers have come full circle and predicted that some long non-coding RNAs can give rise to small proteins that have biological functions. A recent study in *The EMBO Journal* describes how researchers have used ribosome profiling to identify several hundred long non-coding RNAs that may give rise to small peptides.

"We have identified hundreds of open reading frames in the long non-coding RNAs of humans and zebrafish that may give rise to functional proteins using ribosome profiling," says Antonio Giraldez, one of the lead authors of the study and a professor at Yale University School of Medicine in the United States.

Ribosome profiling allows scientists to measure how much RNA is translated into protein. The method allows direct quantification of the messenger RNA fragments protected by the ribosome after digestion with the enzyme nuclease. The nucleases destroy the bonds between the exposed nucleotides that make up RNA and which are not protected by the protein-making machinery of the ribosome. What is left behind is a measurable amount of RNA destined to produce protein.

The researchers were able to visualize translation and the movement of the ribosome every three nucleotides, which corresponds to the size of

each codon on the RNA producing an amino acid. This was possible by combining the high resolution of ribosome profiling with a bioinformatic tool developed in the Giraldez laboratory called ORFScore.

"Crucial to our study was the parallel use of a second computational method that relies on a bioinformatic tool called micPDP," says Giraldez. "micPDP revealed that the RNAs identified by ribosome profiling correspond to peptides that have been conserved over the course of evolution. This strongly suggests that these genes encode proteins that have specific functions in these animals."

As a further validation of their method, the scientists went one step further and used mass spectrometry to detect and characterize almost 100 of the peptides coded by the RNAs.

Until recently, long non-coding RNAs were thought to be restricted to the more mundane but nonetheless important structural roles that are essential to support the function of the cell. "We think the main reason that these small functional peptides have been missed in earlier studies is due to the assumptions that have to be made when assigning functions to large numbers of genes," says EMBO Member Nikolaus Rajewsky, Professor at the Max-Delbrück-Center in Berlin, Germany, Director of the Berlin Institute for Medical Systems Biology and one of the lead authors whose team contributed the micPDP computational method to identify conserved micropeptides. "Short open reading frames are so numerous that by design standard genome annotation methods have to filter out short open reading frames."

There are many short peptides in nature, for example neuropeptides or insulin, but unlike the small peptides arising from long non-coding RNAs they are produced as larger preproteins that need to be trimmed to their final size. The first reports of activities for the small peptides

produced by long non-coding RNAs have already begun to emerge. Schier and colleagues recently reported in *Science*[1] a small peptide that functions as a signal to promote cell motility in the early fish embryo. The aptly named Toddler protein arises from long non-coding RNAs and acts as an activator of a G protein coupled receptor, one of the essential signaling molecules in the cell. Earlier work showed that a long non-coding RNA produced by the tarsal-less/polished rice/mille-pattes gene encodes small peptides that control epithelial morphogenesis in Drosophila and the flour beetle Tribolium.

"Our identification of hundreds of translated small open reading frames significantly expands the set of micropeptide-encoding vertebrate genes providing an entry point to investigate their real life functions," says Giraldez.

"The peptide predictions reported in these studies are tantalizing, but this is just the first step. Things should get really interesting as the community explores the functions of the predicted peptides in vivo," says Stephen M. Cohen, Professor at the Institute of Molecular and Cell Biology in Singapore who is not an author of the paper. "I imagine that we'll be hearing a lot about this new peptide world in the years to come."

More information: Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation, Ariel A. Bazzini, Timothy G. Johnstone, Romain Christiano, Sebastian D. Mackowiak, Benedikt Obermayer, Elizabeth S. Fleming,Charles E. Vejnar, Miler T. Lee, Nikolaus Rajewsky, Tobias C. Walther and Antonio J. Giraldez, emboj.embopress.org/content/ea … 04/04/embj.201488411

[1] Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors (2014) Pauli et al. *Science* 14 February 2014: 343 DOI: 10.1126/science.1248636

Provided by European Molecular Biology Organization