miRNAO: An Ontology for microRNAs

Vicky Dritsou [1], Pantelis Topalis [1], Emmanuel Dialynas [1], Elvira Mitraka [1,2] and Christos Louis [1,2*]

[1]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology - Hellas, Greece
[2]Department of Biology, University of Crete, Greece

ABSTRACT

MicroRNAs (miRNAs) are short RNA molecules (∼22 nt long) that act as gene regulators in many eukaryotes. By binding to complementary strands of messenger RNAs (mRNAs), they induce translational repression or target degradation. Since their discovery in 1993 (Lee *et al.*, 1993), they have received a lot of attention in the research area and some thousands of miRNAs have been reported in the literature; version 18.0 of miRBase (Kozomara and Griffiths-Jones, 2011) contains 18.226 entries representing precursor miRNAs in 168 species, which express 21.643 mature miRNA products.

Each of these small RNA structures regulates many target mRNAs (Baek *et al.*, 2008), while the detection of the targets is a great challenge for researchers. Besides the number of identified miRNAs that grows very fast, the experiments performed aiming at detecting targets have also increased. Large data sets of miRNAs and their experimentally detected targets are stored in various databases. Examples of popular miRNA databases are microRNA.org (Betel *et al.*, 2008), miRBase (Kozomara and Griffiths-Jones, 2011) and TarBase 6.0 (Vergoulis *et al.*, 2011).

Sharing knowledge among different knowledge bases is beneficial and can be achieved with the use of ontologies. The classification of the aforementioned huge data sets under an ontology that expresses the features of miRNAs and their functions can facilitate information discovery and extraction, while it can also be used to inter-relate the databases and/or their schemata. Yet, there is no ontology representing the necessary information regarding miRNAs. The only ontology related to miRNAs is the Ontology for MicroRNA Target Prediction (Townsend *et al.*, 2010), which only captures the information required to express the prediction of targets and does not fully cover all the aspects considered in the area. On the other hand, RNA Ontology (RNAO) (Hoehndorf *et al.*, 2011) serves as an upper ontology, while miRNAO has been instead developed as an application ontology.

In an attempt to capture and represent the existing knowledge regarding miRNAs and their functions, we have developed miRNAO. miRNAO has been designed according to the OBO Foundry Principles (OBO-Foundry, 2006), as these are described in the current release. Moreover, the Basic Formal Ontology (BFO) (Simon *et al.*, 2006) has been used for the classification of the upper classes. Since many ontologies of the domain adhere to the conceptualization of BFO, interoperability among them is thus achieved.

The set of concepts captured and described in miRNAO includes among others molecular entities and parts, their structures, the processes they take place in, the functions and their roles within. Some information related to miRNAs has already been expressed in other popular ontologies. A decision we had to take while developing miRNAO was to either borrow the existing terms and their definitions from other ontologies or to create our own concepts and define them regardless of the existing sources. Even though creating our own terms would allow us to define them according only to our scope, we chose not to re-create those. Instead, we prefer to re-use the existing knowledge by importing them from other ontologies, in an attempt to facilitate knowledge sharing and reuse. Interoperability among different databases is then easier achieved. The majority of shared terms come from the Gene Ontology (GO) (Ashburner *et al.*, 2000) and the Sequence Ontology (SO) (Eilbeck *et al.*, 2005).

miRNAO is the first ontology designed to express all the information related to miRNAs. Exploiting this ontology in different knowledge bases can be very beneficial, both by extracting information from each source and also by interchanging and deriving information from different sources.

*To whom correspondence should be addressed: louis[at]imbb.forth.gr

ACKNOWLEDGEMENTS

The work was supported by the Hellenic General Secretariat for Research and Technology (MicroRNA project, 09SYN-13-1055) and, in part, by the VectorBase project (NIAID).

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nat Genet*, **25**(1), 25–29.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, **455**(7209), 64–71.

Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. (2008). The microrna.org resource: targets and expression. *Nucleic Acids Research*, **36**, D149–D153.

Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, **6**(5), R44+.

Hoehndorf, R., Batchelor, C., Bittner, T., Dumontier, M., Eilbeck, K., Knight, R., Mungall, C. J., Richardson, J. S., Stombaugh, J., Westhof, E., Zirbel, C. L., and Leontis, N. B. (2011). The RNA Ontology (RNAO): An ontology for integrating RNA sequence and structure data. *Appl. Ontol.*, **6**(1), 53–89.

Kozomara, A. and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, **39**(suppl 1), D152–D157.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**(5), 843–854.

OBO-Foundry (2006). Available at http://obofoundry.org/crit.shtml.

Simon, J., Santos, M. D., Fielding, J., and Smith, B. (2006). Formal ontology for natural language processing and the integration of biomedical databases. *Int J Med Inform*, **75**(3-4), 224–231.

Townsend, C., Huang, J., Dou, D., Dalvi, S., Hayes, P. J., He, L., chang Lin, W., Liu, H., Rudnick, R., and Shah, H. (2010). OMIT: Domain Ontology and Knowledge Acquisition in MicroRNA Target Prediction. In *OTM Conferences (2)*, pages 1160–1167.

Vergoulis, T., Vlachos, I., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A. (2011). TarBase 6.0: Capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res*.