Functional annotation of human microRNAs using the Gene Ontology

Barbara Kramarz[1], Rachael P. Huntley[1], Shirin C. C. Saverimuttu[1], Hao Chen[1], Alex Ignatchenko[2], Maria J. Martin[2], Rina Bandopadhyay[3], Manuel Mayr[4], Nigel M. Hooper[5], David Brough[5], Ruth C. Lovering[1]

[1]Functional Gene Annotation, Preclinical and Fundamental Science, Institute of Cardiovascular Science, UCL.
[2]European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Hinxton.
[3]UCL Institute of Neurology and Reta Lila Weston Institute of Neurological Studies.
[4]King's British Heart Foundation Centre, King's College London.
[5]Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester.

Background and Objectives

Experimental data describing the regulation of developmental and cellular processes by microRNAs (miRs) need to be optimally organised to allow inclusion of this data in pathway and network analysis tools. As the association of proteins with terms from the Gene Ontology (GO)[1,2] has proven highly effective for large-scale functional analysis of high-throughput (HTP) data[3], our aim has been to apply this approach to annotate human miRs.

Having established GO Consortium guidelines[4,5] for curation of miRs, we have provided >6000 GO annotations for 660 miRs, including information describing over 400 human miRs and their regulation targets. To date we have focused on miRs involved in cardiovascular and dementia-relevant processes. In addition, we have shown how our annotations are contributing to the understanding of the role of miRs in co-ordinating the regulation of specific processes.

Here we provide an overview of GO, explain the miR curation approach and provide examples of miR GO annotations and their uses.

[1] Ashburner et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25-9.
[2]The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330-D338.
[3]Lovering et al. Improving Interpretation of Cardiac Phenotypes and Enhancing Discovery With Expanded Knowledge in the Gene Ontology. Circ Genom Precis Med. 2018; 11(2):e001813.
[4]Huntley et al. Guidelines for the functional annotation of microRNAs using the Gene Ontology. RNA. 2016;22(5):667-76.
[5]Huntley et al. Expanding the horizons of microRNA bioinformatics. RNA. 2018;24(8):1005-1017.

Gene Ontology: a dictionary for biology

- Gene Ontology (GO): a collaborative effort to provide freely-available, standardised descriptions of proteins, noncoding RNAs and complexes across all species and biological fields.
- GO comprises three types of structured controlled vocabularies that describe gene products, such as proteins and miRs, in terms of their associated *molecular functions*, *biological processes*, and localisation to *cellular components*.
- Originally developed in 1998, GO has grown to include 45,000 terms, arranged as a hierarchy, describing a wide range of concepts, and linked by different types of relationships (Figure 2).
- GO is an essential resource for analysis of high-throughput data, facilitating the grouping of genes into common pathways, functions and cellular locations.

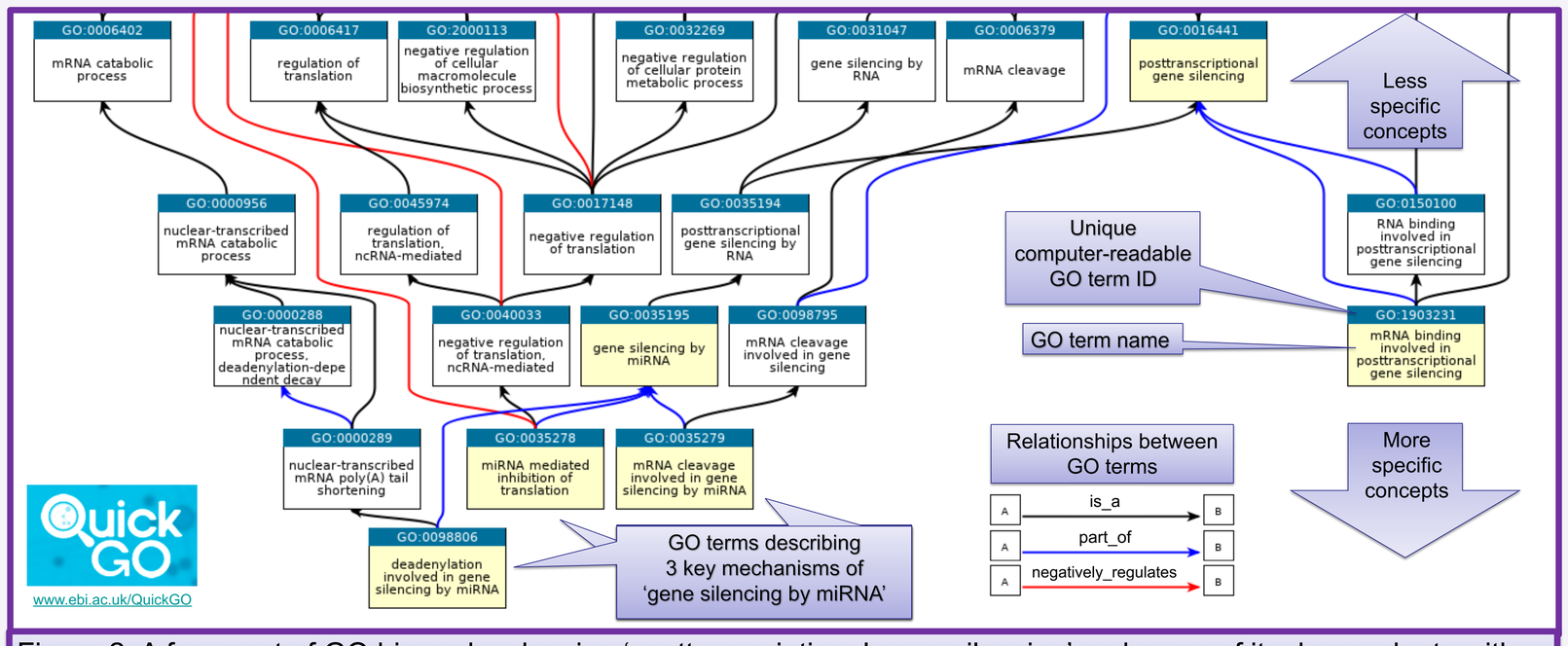

Figure 2. A fragment of GO hierarchy showing 'posttranscriptional gene silencing' and some of its descendants with miR-specific terms shown on yellow background. Image adapted from QuickGO.

GO annotation of microRNAs

Specific unique criteria need to be fulfilled in order to capture biological roles of miRs.

1) Experimental evidence of miR-mRNA binding is essential for annotating the interaction between a miR and its target (Figure 1).
2) Sequence and species of miR used in the described experiments needs to be provided either directly within the research article, or indirectly in a manufacturer's catalogue.

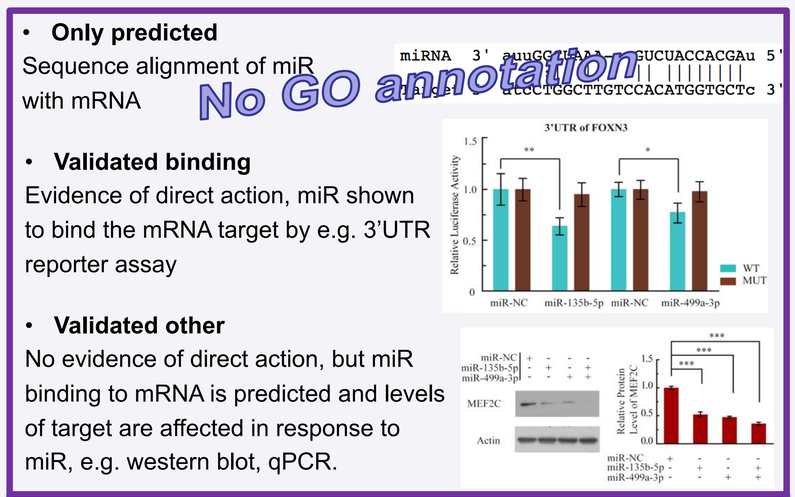

Figure 1. Guidelines for GO annotation of miR-target interaction. Images adapted from Xu *et al*. Sci Rep. Jul 2015. PMID:26184978. Open Access.

Entity identifier and name	GO term identifier and name	Evidence	Reference	Taxon	Curator	Extension
RNAcentral:URS000013D17D_9606 (hsa-miR-19b-3p)	GO:1903231 (F) mRNA binding involved in posttranscriptional gene silencing	ECO:0000314 IDA	PMID:25084135	9606 Homo sapiens	BHF-UCL	has_direct_input (ENSEMBL:ENSG00000165029) (*ABCA1*)
RNAcentral:URS000013D17D_9606 (hsa-miR-19b-3p)	GO:0035278 (P) miRNA mediated inhibition of translation	ECO:0000314 IDA	PMID:22105995	9606 Homo sapiens	ARUK-UCL	regulates_translation_of (ENSEMBL:ENSG00000137462) more... (*TLR2*)
RNAcentral:URS000013D17D_9606 (hsa-miR-19b-3p)	GO:0048471 (C) perinuclear region of cytoplasm	ECO:0000250 ISS	GO_REF:0000024	9606 Homo sapiens	BHF-UCL	part_of (CL:0000746) (cardiac muscle cell)
RNAcentral:URS000013D17D_9606 (hsa-miR-19b-3p)	GO:1900165 (P) negative regulation of interleukin-6 secretion	ECO:0000314 IDA	PMID:22105995	9606 Homo sapiens	ARUK-UCL	

Figure 3. Examples of GO annotations contributed by BHF-UCL and ARUK-UCL. Key: F, function; P, process; C, cellular component; IDA, inferred from direct assay; ISS, inferred from sequence similarity. Images adapted from QuickGO.

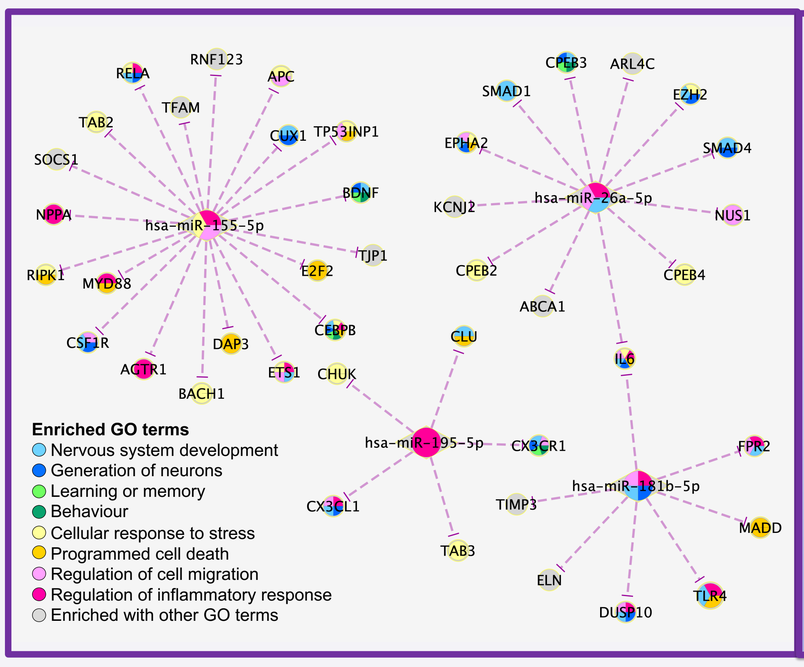

Figure 4. Examples of miR-target molecular interaction networks. The network was seeded with four of the miR RNAcentral identifiers, annotated as a part of the ARUK-UCL GO annotation initiative. The seed miRs are in the centre of the four hubs. Dashed purple edges represent experimentally demonstrated associations between miRs and their mRNA targets, represented by nodes labelled with HGNC[7]-approved gene symbols; the cap on the purple edge faces the target of the miR regulation. Node colours correspond to GO terms enriched (overrepresented) within the group of gene products shown in the figure and are explained in the 'Enriched GO terms' key. The networks were constructed in Cytoscape 3.7.1[6] using molecular interaction data from the EBI-GOA-miR file at PSICQUIC (1st July 2019).

[6] Shannon et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11):2498-504.
[7] Gray et al. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 2015; 43(Database issue):D1079-85.

Conclusions and Future Prospects

Gene Ontology (GO, GENEONTOLOGY Unifying Biology) is one of the major resources used for analyses and interpretation of biomedical 'big data'. It has been our goal to develop GO guidelines for organising the knowledge and information about the biological roles of miRs by capturing, firstly, the regulation of expression of target genes, and, secondly, the effects of this regulation on the occurrences and/or outcomes of specific biological processes. Our focused and systematic approach to GO curation of miRs has so far resulted in >6000 GO annotations for 660 miRs, capturing the mRNA targets for 415 of these miRs. This includes >4500 annotations specifically for 442 human miRs, capturing the mRNA targets for 254 of these miRs (QuickGO, EBI statistics: 31st August 2019). The miR annotations are freely available in the QuickGO and AmiGO 2 browsers, RNAcentral, miRbase, Ensembl and the PSICQUIC web server, enabling our data to be included in many functional analysis tools and used to interpret large datasets from HTP studies. Our plan over the next few years is to contribute to the advancement of miR research by continuing to build a resource comprising of high-quality, reliable functional annotations for human miRs.

Alzheimer's Research UK
British Heart Foundation
Annotation initiative funded by Alzheimer's Research UK grants ARUK-NAS2017A-1, ARUK-NSG2018-003 and British Heart Foundation grant RG/13/5/30112. The Functional Gene Annotation team is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The UCL curators are members of the GO Consortium.

UCL Functional Gene Annotation
GENEONTOLOGY Unifying Biology www.geneontology.org
www.ucl.ac.uk/functional-gene-annotation/

r.lovering@ucl.ac.uk
barbara.kramarz@ucl.ac.uk
@UCLgene

MANCHESTER 1824 The University of Manchester
EMBL-EBI