Bioinformatics, 31(14), 2015, 2403–2405
doi: 10.1093/bioinformatics/btv140
Advance Access Publication Date: 10 March 2015
Applications Note

OXFORD

Systems biology

miR2GO: comparative functional analysis for microRNAs

Anindya Bhattacharya[1,2] and Yan Cui[1,2,*]

[1]Department of Microbiology, Immunology and Biochemistry and [2]Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, TN 38163, USA

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on June 30, 2014; revised on February 25, 2015; accepted on March 5, 2015

Abstract

Summary: miR2GO is a web-based platform for comparative analyses of human miRNA functions. It includes two programs: miRmut2GO and miRpair2GO. miRmut2GO implements a knowledge-based method to assess the functional effects of genetic and somatic mutations in microRNA seed regions. The functional effects of a mutation are analysed by semantic comparison of enriched gene ontology (GO) annotations of the target gene sets for the wild-type and mutated alleles. miRpair2GO compares the functions of two different miRNAs based on the enriched functional annotations of their target gene sets.

Availability and implementation: The miR2GO web server is available at http://compbio.uthsc.edu/miR2GO.

Contact: ycui2@uthsc.edu

1 Introduction

MicroRNAs (miRNA) are important post-transcriptional regulators of gene expression. It has been estimated that ~60% of protein coding genes are potential targets of miRNA regulation (Friedman *et al.*, 2009). The most important regulatory signal of a miRNA, the information guiding its target recognition, is primarily encoded in the seed region (nucleotides 2–7 in mature miRNA sequences). Because a miRNA often have a large number of target genes, a miRNA seed mutation, may affect the regulation of many genes simultaneously. Thus, genetic and somatic mutations in miRNA seeds may have broad impacts on target gene expression and have been linked to various diseases (Bhattacharya *et al.*, 2013, 2014; Iliff *et al.*, 2012; Mencia *et al.*, 2009).

MicroRNA functions can be predicted and annotated by linking the target genes to diseases and gene ontology (GO) terms (Lagana *et al.*, 2009; Nam *et al.*, 2008). However, such analyses are based only on wild-type miRNAs and do not study the effects of miRNA mutations. The functional effects of a miRNA seed mutation can be assessed by comparative analysis of predicted target genes for the wild-type and mutated alleles of the miRNA. Differences and similarities in the enriched functional annotations of the reference (In this paper, the set of target genes for the wild-type miRNA are called 'reference targets', the set of target genes for the mutated miRNA are called 'derived targets', the genes in both sets are called 'common targets'.) and derived target genes reflect the overall functional impacts of the mutation. Here, we developed the miR2GO web server to provide a computational tool for the assessment of functional impacts of genetic and somatic mutations in miRNA seed regions. The first program of this web server, miRmut2GO, allows users to analyse the changes of target genes caused by miRNA seed mutations and view the functional impacts of these changes in GO graphs. The second function of the web server, miRpair2GO, allows users to perform comparative functional analysis of different miRNAs.

2 Tool description

2.1 miRmut2GO

miRmut2GO is a web-based program for evaluating the functional effects of mutations in miRNA seeds. First, miRmut2GO predicts the reference and derived target genes for the miRNAs with

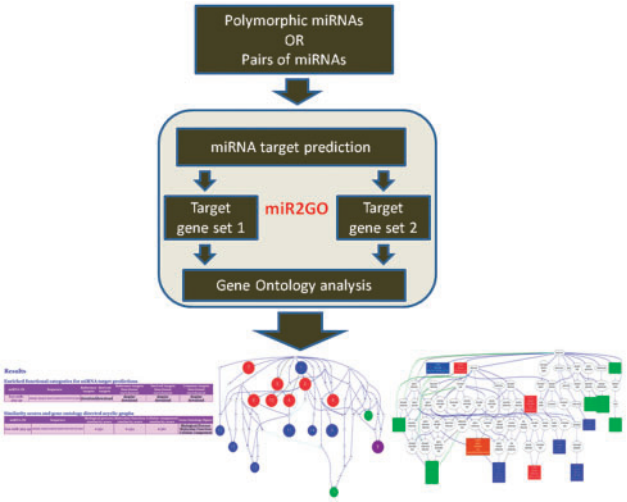

Fig. 1. A summary of miR2GO functions

seed mutations. Users can choose from two frequently used miRNA target prediction algorithms, TargetScan (Lewis *et al.*, 2005) and miRanda (Enright *et al.*, 2003). Users can also combine the prediction results from the two algorithms by selecting the union or intersection options. Second, miRmut2GO identifies the enriched functional annotations for the reference and derived target genes. The enriched functional annotations are obtained by using gProfileR, an R interface for the web-based GO tool gProfiler (Reimand *et al.*, 2011). Third, a semantic similarity score (Wang *et al.*, 2007) for the two sets of enriched functional annotations is calculated using GOSemSim (Yu *et al.*, 2010), an R library for semantic comparison between two groups of GO terms. The similarity score varies from 0 to 1, where 0 indicates no similarity and 1 indicates complete similarity. Fourth, miRmut2GO prepares GO graphs to visualize the functional effects of miRNA seed mutations. RamiGO (Schroder *et al.*, 2013), an R/Bioconductor library for AmiGO GO visualization, and Graphviz (Gansner and North, 2000) are used for preparing GO graphs. Figure 1 shows the functional diagram for miRmut2GO. In the GO graph, the enriched GO terms are highlighted. The color of a node is determined by three numbers—the number of reference targets, the number of derived targets and the number of common targets, which are annotated to this GO term. These three numbers are used to determine the blue, green and red components of the node color. The node size is determined by the total number of target genes annotated to this GO term. Such GO graphs provide a global view of the functional impacts of the miRNA seed mutation. A more detailed GO graph with the nodes labeled by GO terms, number of genes and *P*-values for functional enrichments is also available from miRmut2GO.

2.2 miRpair2GO

Different miRNAs perform different functions by binding to different sets of target genes. Therefore, we can compare the functions of two miRNAs by comparing the functional annotations of their target genes. miRpair2GO is a tool for comparative functional analysis of different miRNAs. The workflow is almost same as that of miRpair2GO except that miRpair2GO takes pairs of different miRNAs as input (Fig. 1).

3 Usage

Both miRmut2GO and miRpair2GO require selection of prediction method, parameter setting for functional enrichment analysis and selection of input data. Options are provided for users to select of either TargetScan or miRanda or combining both methods for miRNA target prediction. Parameter settings for GO enrichment analyses include specification of a *P*-value threshold for functional enrichment and selection of a hierarchical filtering level. The hierarchical filtering is used for grouping the similar GO terms in the hierarchy of GO graph. After setting parameters, users need to specify miRNA IDs or upload miRNA sequence data. For miRmut2GO users also need to specify the mutations in miRNA seeds. miRmut2GO supports the search of multiple single nucleotide polymorphisms (SNPs). Similarly, miRpair2GO supports multiple miRNA pairs as input. After the query is submitted, miR2GO will run the workflows described earlier. Upon finishing all the calculations, the programs first give tabular results with functional similarity scores. The lists of predicted target genes and enriched functional annotations and GO figures can be displayed by clicking the links in the table. A detailed description of the usage of miRmut2GO and miRpair2GO can be found in the help page of miR2GO (http://compbio.uthsc.edu/miR2GO/help.php).

4 Discussion

With the advance in genomic technologies, the number of newly identified genetic and somatic mutations in miRNA seeds has been increasing very rapidly in recent years. There was virtually no known miRNA seed SNPs when we created the PolymiRTS database (Bao *et al.*, 2007) and only 20 SNPs were mapped to the human miRNA seed regions when we developed PolymiRTS 2.0 (Ziebarth *et al.*, 2012). The number of human miRNA seed SNPs increased to 271 in PolymiRTS 3.0 (Bhattacharya *et al.*, 2014). Currently, 517 SNPs can be mapped to the seed regions of 414 human miRNAs using the genomic coordinates of SNPs in dbSNP (Sherry *et al.*, 2001) and the miRNA data in miRBase (Kozomara and Griffiths-Jones, 2014). Somatic mutations have also been found in miRNA seed regions (Bhattacharya *et al.*, 2013; Silva *et al.*, 2013). There is a great need to analyse the functional impacts of the miRNA seed mutations. miR2GO provides a web-based tool to address this challenge. It integrates miRNA target prediction, GO enrichment analysis and semantic comparison of functional annotations to analyse the functional impacts of mutations in miRNA seed regions. It also uses the same workflow to compare the functions of different miRNAs. It is noteworthy that miR2GO can also be used to predict the functional impacts of any hypothetical seed mutations or compare the functions of designed miRNA sequence with existing miRNAs, which might be useful in artificial miRNA design (Tay *et al.*, 2015).

miR2GO relies on miRNA target prediction tools to generate target gene sets. The default miRNA target prediction tool of miR2GO is TargetScan. In a recent comprehensive assessments of miRNA target prediction methods (Fan and Kurgan, 2015), TargetScan was one of the best predictors with an AUC (area under the ROC curve) of 0.748 at the gene level. Although this is a reasonably good predictive performance, obviously miR2GO can benefit from more accurate miRNA target prediction. Because miR2GO uses a modular framework, better target prediction tools can be easily incorporated when they become available.

Funding

This work was partly supported by The University of Tennessee Center for Integrative and Translational Genomics.

Conflict of Interest: none declared.

References

Bao,L. *et al.* (2007) PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids* Res., **35**, D51–D54.

Bhattacharya,A. *et al.* (2013) SomamiR: a database for somatic mutations impacting microRNA function in cancer. *Nucleic Acids Res.*, **41**, D977–D982.

Bhattacharya,A. *et al.* (2014) PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.*, **42**, D86–D91.

Enright,A.J. *et al.* (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.

Fan,X. and Kurgan,L. (2015) Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Briefings Bioinform*, (in press) doi:10.1093/bib/bbu044.

Friedman,R.C. *et al.* (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Gansner,E.R. and North,S.C. (2000) An open graph visualization system and its applications to software engineering. *Softw. Pract. Exper.*, **30**, 1203–1233.

Iliff,B.W. *et al.* (2012) A single-base substitution in the seed region of miR-184 causes EDICT syndrome. *Invest. Ophthalmol. Vis. Sci.*, **53**, 348–353.

Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.

Lagana,A. *et al.* (2009) miRo: a miRNA knowledge base. *Database*, **2009**, bap008.

Lewis,B.P. *et al.* (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Mencia,A. *et al.* (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat. Genet.*, **41**, 609–613.

Nam,S. *et al.* (2008) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.

Reimand,J. *et al.* (2011) g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.*, **39**, W307–W315.

Schroder,M.S. *et al.* (2013) RamiGO: an R/Bioconductor package providing an AmiGO visualize interface. *Bioinformatics*, **29**, 666–668.

Sherry,S.T. *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.

Silva,J. *et al.* (2013) Dysregulation and recurrent mutation of miRNA-142 in de novo AML. *Blood*, **122**, 472–472.

Tay,F.C. *et al.* (2015) Using artificial microRNA sponges to achieve microRNA loss-of-function in cancer cells. *Adv. Drug Deliv. Rev.*, **81**, 117–127.

Wang,J.Z. *et al.* (2007) A new method to measure the semantic similarity of GO terms. *Bioinformatics*, **23**, 1274–1281.

Yu,G. *et al.* (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, **26**, 976–978.

Ziebarth,J.D. *et al.* (2012) PolymiRTS database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.*, **40**, D216–D221.