



APPLICATION NOTE

C2Analyzer: Co-target–Co-function Analyzer

Md. Aftabuddin ¹, Chittabrata Mal ², Arindam Deb ², Sudip Kundu ^{2,*}¹ West Bengal University of Technology, Kolkata 700064, India² Department of Biophysics, Molecular Biology & Bioinformatics, University of Calcutta, Kolkata 700009, India

Received 23 October 2013; revised 25 March 2014; accepted 26 March 2014

Available online 24 May 2014

Handled by Andreas Keller

KEYWORDS

MicroRNA;
 Post-transcriptional
 regulation;
 Gene ontology;
 Co-target;
 Co-function

Abstract MicroRNAs (miRNAs) interact with their target mRNAs and regulate biological processes at post-transcriptional level. While one miRNA can target many mRNAs, a single mRNA can also be targeted by a set of miRNAs. The targeted mRNAs may be involved in different biological processes that are described by gene ontology (GO) terms. The major challenges involved in analyzing these multitude regulations include identification of the combinatorial regulation of miRNAs as well as determination of the co-functionally-enriched miRNA pairs. The C2Analyzer: Co-target–Co-function Analyzer, is a Perl-based, versatile and user-friendly web tool with online instructions. Based on the hypergeometric analysis, this novel tool can determine whether given pairs of miRNAs are co-functionally enriched. For a given set of GO term(s), it can also identify the set of miRNAs whose targets are enriched in the given GO term(s). Moreover, C2Analyzer can also identify the co-targeting miRNA pairs, their targets and GO processes, which they are involved in. The miRNA–miRNA co-functional relationship can also be saved as a .txt file, which can be used to further visualize the co-functional network by using other software like Cytoscape. C2Analyzer is freely available at www.bioinformatics.org/c2analyzer.

Introduction

MicroRNAs (miRNAs), a class of 20–24 nucleotide small non-coding RNAs, play important roles in post-transcriptional gene regulation by targeting mRNAs for cleavage or repressing

translation [1]. Although thousands of miRNAs have been identified, the function of most miRNAs involved in biological networks remains unclear [2].

The miRNAs regulate a wide range of cellular functions, such as development, differentiation, proliferation, apoptosis and metabolism. It has been hypothesized that a single miRNA can concurrently down-regulate hundreds of target mRNAs [3] and thus affect the expression of numerous genes at one time. Functionally related genes are usually regulated by a group of miRNAs instead of one miRNA. The down-regulation of a candidate target by a set of miRNAs depends on how many and which miRNAs are expressed in the cell at a given time. However, recent studies indicate that a relatively small number of miRNAs can set up remarkably complex

* Corresponding author.

E-mail: skbmbg@caluniv.ac.in (Kundu S).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.



spatial and temporal patterns of gene expression by means of combinatorial gene regulation [4,5], *i.e.*, regulation of a gene by two or more miRNAs simultaneously.

This combinatorial regulation of miRNAs in any species is common and studying the potential functional effect of such regulation is an interesting and challenging area of miRNA research. However, the existing tools for analysis of the combinatorial regulation have limitations. For example, while GeneSet2miRNA [6] can predict a group of miRNAs regulating a significant subset of genes, the prediction of statistically enriched biological functions regulated by these miRNAs is missing in this tool. On the other hand, CORNA [7] can be used to identify the significant mRNA sets as well as gene ontology (GO) terms regulated by a single miRNA. However, it is not possible to get a list of significant mRNAs or GO terms that are co-targeted by a pair of miRNAs using CORNA.

It has been demonstrated that the analysis of co-targeting and co-functional miRNA pairs can generate the functional modules, which can further establish the synergistic behavior of miRNAs [8]. Here, synergistic effect is considered when at least one functional module (*i.e.*, GO) is significantly co-targeted by a pair of miRNAs. For a given set of input data (miRNA–mRNA and mRNA–GO process), our recently-developed C2Analyzer reported in the present study can identify (i) the statistically significant co-functional miRNA pairs, their targets and GO processes involved and (ii) a candidate set of miRNAs whose targets are enriched in the given GO terms. In addition, users can also obtain some straightforward results showing the one-to-many and many-to-many relationships among the miRNA–mRNA–GO terms.

Methods

C2Analyzer requires two files as input: one mapping the miRNAs to the target genes and the other mapping each gene to the GO IDs that it belongs to. The tool can perform five types of analysis starting from these two files (see [Figure 1](#) and output section). The main objectives are to (i) identify the targets shared by two miRNAs and find the GO terms that are enriched in these targets and (ii) identify the miRNAs with enriched targets associated with a set of GO terms. In addition, the tool can also be used to (iii) identify the mRNAs and their GO terms associated with a list of miRNAs, (iv) identify the miRNAs targeting a given set of

mRNAs and (v) identify the mRNAs that are co-targeted by a given set of miRNAs.

Identification of co-functionally significant miRNA pairs

We have identified each miRNA pair and their co-targeting mRNAs as a target subset and candidate functional modules in the target subset by performing functional enrichment in each biological process category. For a given miRNA pair of M1 and M2, we identify the target subset they co-target ($M1 \cap M2$). The subset must contain a minimum number of target genes (*i.e.*, O_{\min}). Here, the value of O_{\min} is set to 3 [8]. First, biological processes where the target subset is enriched are identified by hypergeometric distribution. The probability PG_i for $M1 \cap M2$ in the GO term ' t ' is calculated according to the equation used in [8],

$$PG_i = 1 - F(n|S, T_i, L) = 1 - \sum_{t=0}^n \frac{\binom{T_i}{t} \binom{S-T_i}{L-t}}{\binom{S}{L}}$$

where S is the number of all targets (default background distribution), T_i is the total number of genes that are annotated in the GO term ' t ' and targeted by miRNAs, L is the size of $M1 \cap M2$, n is the number of targets in $M1 \cap M2$ that are also annotated to term ' t '. Here, $i = 1, 2, \dots, I$ and I is the total number of GO terms that we considered. At the given significance level, we can obtain not only the set with enriched function terms but also capture the set GO_{M1M2} with the subsets in $M1 \cap M2$ that are annotated to each term in the previous set. Namely, GO_{M1M2} is the set of candidate functional modules. Next, the functional modules in GO_{M1M2} are identified. After performing the function enrichment, M1 and M2 are considered to be synergistic if they co-target at least one functional module.

Identification of the miRNAs associated with a set of GO terms

To identify the miRNAs whose targets are enriched in the given GO ID(s), the chi-square test is performed.

System implementation

This web-based tool is supported on any web-browser at any supported resolution. Uploading the data may take some time

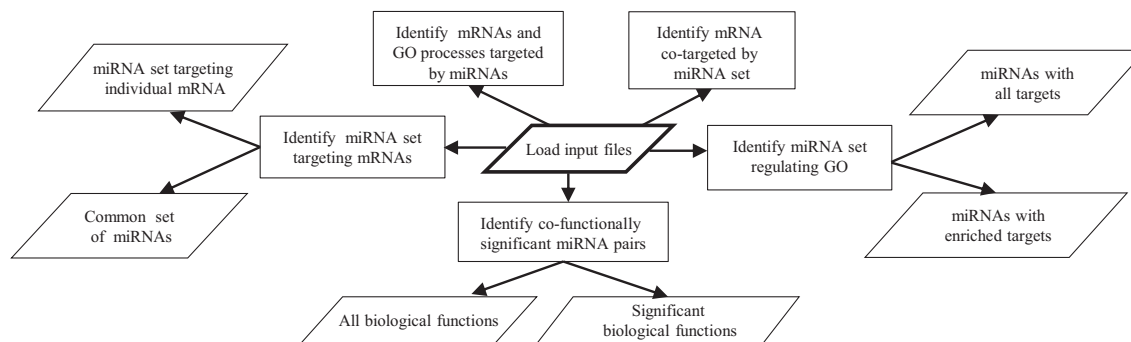


Figure 1 Outline of functionality of C2Analyzer

The arrows indicate direction and branches of analysis with input files as the start and the leaves as the last stages (save results) in the analysis.

and analysis time depends on the size of input files. This tool is developed using Perl CGI and is hosted on an Apache web server. The web pages are designed with HTML.

Web interface

Input

User can upload their own data as two relation files (two column data separated by tab). One is miRNA *vs.* target mRNA, and the other is target mRNA *vs.* GO process. Test sample datasets are provided for the users. Users can create these files by following the instructions given in the online documentation section.

Output

The miRNA–miRNA co-functional relationship file can be saved as a .txt file, which can be further utilized to visualize the co-functional network in Cytoscape like software. Output for other types of analysis can be saved as .csv format. The following types of analysis could be done by C2Analyzer.

mRNA co-targeted by miRNA set

With a given set of miRNAs, users will get a set of mRNAs co-targeted by all of the input miRNAs.

Co-functionally significant miRNA pair

Here, users can get two types of output with a given set of miRNAs. First, users can get a set of GO terms that are regulated by miRNA pairs. Second, the tool can also perform hypergeometric analysis to identify set of GO(s) with enriched targets regulated by miRNA pairs. Users can set the *P* value (the default value is 0.05).

miRNA set regulating GO

A user can determine a set of miRNAs that are involved in a single GO process. Further, the tool can also identify a set of miRNAs that regulate a set of GO processes with enriched targets. For the latter case, the chi-square test is performed and the user can choose the *P* value cut-off.

mRNA and GO targeted by miRNA

For a single miRNA, the output can generate two tables: one describing a list of mRNAs and the other showing a list of GO IDs targeted by this miRNA.

miRNA set targeting mRNAs

For a given set of mRNAs, the user can get the individual lists of miRNAs targeting each of these mRNAs. The user can also obtain a common list of miRNAs co-targeting all of the mRNAs given as input.

Case study

Here we present an example to illustrate the use of C2Analyzer to identify the co-functionally significant miRNA pairs. A subset of cadmium induced-miRNAs in the rice plant collected from literature [9] was used for this analysis. We took four miRNAs, namely osa-miR156l, osa-miR156k, osa-miR156a

and osa-miR162a, which were reported to be down-regulated due to cadmium stress [9].

C2Analyzer is able to find target mRNAs and a list of GOs regulated by all possible pairwise combinations of these four miRNAs. Moreover, based on hypergeometric distribution analysis, the tool is also able to identify a list of significantly-enriched GOs for every possible pair of miRNAs. For example, the miRNA pair osa-miR156a/k can co-target 11 mRNAs (Table 1) and thus can co-regulate a total number of 13 biological functions (GO). While C2Analyzer is used to identify the GOs in which the miRNA targets are significantly enriched ($P < 0.05$), we get only 4 GO processes (Table 2). The four GO processes identified are GO:0003677 (DNA binding, molecular function), GO:0008270 (zinc ion binding, molecular function), GO:0046872 (metal ion binding, molecular function)

Table 1 Co-targeted mRNAs by miRNA pair osa-miR156a/k and the related GOs

Common mRNA set	GO
LOC_Os01g69830.1	GO:0003677, GO:0005634
LOC_Os02g04680.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os02g04680.2	GO:0003677, GO:0005634
LOC_Os02g07780.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os02g07780.2	GO:0003677, GO:0005634
LOC_Os06g45310.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os07g32170.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os08g39890.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os08g41940.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872
LOC_Os09g31438.1	GO:0003677, GO:0005634, GO:0008270, GO:0046872

Table 2 Significantly-enriched GOs and mRNA sets co-targeted by miRNA pair osa-miR156a/k

GO	mRNA set
GO:0003677	LOC_Os01g69830.1, LOC_Os02g04680.1, LOC_Os02g04680.2, LOC_Os02g07780.1, LOC_Os02g07780.2, LOC_Os06g45310.1, LOC_Os07g32170.1, LOC_Os08g39890.1, LOC_Os08g41940.1, LOC_Os09g31438.1
GO:0005634	LOC_Os01g69830.1, LOC_Os02g04680.1, LOC_Os02g04680.2, LOC_Os02g07780.1, LOC_Os02g07780.2, LOC_Os06g45310.1, LOC_Os07g32170.1, LOC_Os08g39890.1, LOC_Os08g41940.1, LOC_Os09g31438.1
GO:0008270	LOC_Os02g04680.1, LOC_Os02g07780.1, LOC_Os06g45310.1, LOC_Os07g32170.1, LOC_Os08g39890.1, LOC_Os08g41940.1, LOC_Os09g31438.1
GO:0046872	LOC_Os02g04680.1, LOC_Os02g07780.1, LOC_Os06g45310.1, LOC_Os07g32170.1, LOC_Os08g39890.1, LOC_Os08g41940.1, LOC_Os09g31438.1

Note: Significantly-enriched GOs were determined using hypergeometric test ($P < 0.05$).

and GO:0005634 (nucleus, cellular component). The enriched targets (LOC_Os01g69830.1, LOC_Os02g07780.1, *etc.*) within these four GOs are annotated as SBP-box genes. It is of note that SBP-box genes encode transcription factors, and GO:0003677 has DNA binding function. It is also known that the SBP-box proteins regulate some stress-responsive genes and thus play important roles in maintaining cellular homeostasis [10]. The cadmium is a metal ion with strong affinity for oxygen, nitrogen and sulfur. Thus, it can inhibit the enzymatic activity by directly blocking protein functions or altering the natural metal centers [11]. To maintain the cellular homeostasis, the cell may overproduce the genes having metal ion binding functionality. Since expression of miRNAs osa-miR156a and osa-miR156k was down-regulated due to cadmium stress [9], their targets, *i.e.*, the SBP-box genes, must have higher expression presumably. Interestingly, one of the enriched molecular functions of the SBP-box genes as annotated in the GO process is metal ion binding (GO:0046872). Thus, this combinatorial regulation may be one of the mechanisms to maintain the cellular homeostasis under cadmium stress.

Future development

We plan to upgrade the current web-based tool as well as to incorporate several existing datasets. For example, in the future, condition-specific miRNAs and gene expression data will be included. The necessary tools for analyzing expression data will also be added to analyze and integrate the expression dataset.

Conclusion

Here we have developed a user-friendly web-tool, C2Analyzer, that uses two input files (miRNA–mRNA and mRNA–GO) to identify (i) co-functionally significant miRNA pair, (ii) miRNA set regulating GO processes with enriched targets, (iii) mRNA and GO targeted by miRNA, (iv) miRNA set targeting mRNA and (v) mRNA co-targeted by miRNA set. C2Analyzer can also be used to analyze datasets for any organism. Our results also indicate that C2Analyzer may help elucidate the functions of some functionally-unidentified miRNAs. As the number of newly-identified miRNAs is growing day by day, the web tool will be helpful for their functional analysis. Finally, we want to mention that this web-based tool is flexible to analyze the general relationship among X , Y and Z with two given input files describing X – Y and Y – Z relationships. For example, if the user provides miRNA–gene and gene–disease files, this tool can find the relationship between miRNA and disease as well.

Authors' contributions

CM and SK conceived the project and prepared the manuscript. MA and CM designed the tool and performed majority of the analysis. CM and AD collected the dataset and

participated in data analysis. All of the authors have read and approved the final manuscript.

Competing interests

The authors declare that no competing interests exist.

Acknowledgements

CM and AD gratefully acknowledge University Grants Commission (Research Fellowship for Meritorious Students in Science), Government of India for financial support and Distributed Information Center, University of Calcutta for computational facilities. SK acknowledges Centre of Excellence on Systems Biology and Biomedical Engineering for partial support. All the authors are grateful to <http://www.bioinformatics.org> for allowing them to host the C2Analyzer tool on their website and also thankful to the anonymous reviewers for their useful suggestions.

References

- [1] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
- [2] Tsang JS, Ebert MS, Oudenaarden A. Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol Cell* 2010;38:140–53.
- [3] Friedman Y, Balaga O, Linial M. Working together: combinatorial regulation by microRNAs. *Adv Exp Med Biol* 2013;774:317–37.
- [4] Joung JG, Fei Z. Identification of microRNA regulatory modules in Arabidopsis via a probabilistic graphical model. *Bioinformatics* 2009;25:387–93.
- [5] Balaga O, Friedman Y, Linial M. Toward a combinatorial nature of microRNA regulation in human cells. *Nucleic Acids Res* 2012;40:9404–16.
- [6] Antonov AV, Sabine D, Philip W, Dominik L, Hans WM. GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res* 2009;37:W323–8.
- [7] Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. *Bioinformatics* 2009;25:832–3.
- [8] Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, et al. MiRNA–miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res* 2011;39:825–36.
- [9] Ding Y, Chen Z, Zhu C. Microarray-based analysis of cadmium-responsive microRNAs in rice (*Oryza sativa*). *J Exp Bot* 2011;62:3563–73.
- [10] Wang Y, Hu Z, Yang Y, Chen X, Chen G. Function annotation of an SBP-box gene in *Arabidopsis* based on analysis of co-expression networks and promoters. *Int J Mol Sci* 2009;10:116–32.
- [11] Tan YF, O'Toole N, Taylor NL, Millar AH. Divalent metal ions in plant mitochondria and their role in interactions with proteins and oxidative stress-induced damage to respiratory function. *Plant Physiol* 2010;152:747–61.