SOFTWARE TOOL ARTICLE

REVISED MIREyA: a computational approach to detect miRNA-directed gene activation [version 2; peer review: 1 approved, 2 approved with reservations]

Anna Elizarova [ID][1,2], Mumin Ozturk[3,4], Reto Guler[3-5], Yulia A. Medvedeva[1,2]

[1]Group of Regulatory Transcriptomics and Epigenomics, Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, 117312, Russian Federation
[2]Department of Biological and Medical Physics, Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, 141701, Russian Federation
[3]International Centre for Genetic Engineering and Biotechnology, Cape Town, Cape Town, 7925, South Africa
[4]Department of Pathology, University of Cape Town, Institute of Infectious Diseases and Molecular Medicine (IDM), Division of Immunology and South African Medical Research Council (SAMRC) Immunology of Infectious Diseases, Faculty of Health Sciences, Cape Town, 7925, South Africa
[5]Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, 7925, South Africa

Abstract

Emerging studies demonstrate the ability of microRNAs (miRNAs) to activate genes via different mechanisms. Specifically, miRNAs may trigger an enhancer promoting chromatin remodelling in the enhancer region, thus activating the enhancer and its target genes. Here we present MIREyA, a pipeline developed to predict such miRNA-gene-enhancer trios based on an expression dataset which obviates the need to write custom scripts. We applied our pipeline to primary murine macrophages infected by *Mycobacterium tuberculosis* (HN878 strain) and detected Mir22, Mir221, Mir222, Mir155 and Mir1956, which could up-regulate genes related to immune responses. We believe that MIREyA is a useful tool for detecting putative miRNA-directed gene activation cases. MIREyA is available from:
https://github.com/veania/MIREyA

Keywords
microRNA, miRNA, non-coding RNA, enhancer, regulator, macrophage, tuberculosis

Open Peer Review

Reviewer Status ? ? ✔

	Invited Reviewers		
	1	**2**	**3**
version 2 (revision) 26 Aug 2021			✔ report
version 1 29 Mar 2021	? report	? report	? report

1. **Michiel Jan Laurens De Hoon** [ID], RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

2. **Xiangfu Zhong** [ID], Karolinska Institutet, Stockholm, Sweden

3. **Chaogang Shao**, Huzhou University, Huzhou, China

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding authors: Anna Elizarova (elizarova@phystech.edu), Yulia A. Medvedeva (ju.medvedeva@gmail.com)

Author roles: Elizarova A: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ozturk M**: Funding Acquisition, Methodology, Writing – Review & Editing; **Guler R**: Funding Acquisition, Writing – Review & Editing; **Medvedeva YA**: Conceptualization, Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

How to cite this article: Elizarova A, Ozturk M, Guler R and Medvedeva YA. **MIREyA: a computational approach to detect miRNA-directed gene activation [version 2; peer review: 1 approved, 2 approved with reservations]** F1000Research 2021, **10**:249 https://doi.org/10.12688/f1000research.28142.2

First published: 29 Mar 2021, **10**:249 https://doi.org/10.12688/f1000research.28142.1

> **REVISED** **Amendments from Version 1**
>
> This version contains several minor changes. First of all, we added a more detailed explanation of arguments which a user provides to our pipeline and description of the output of the pipeline. Secondly, the missing number of enhancers is now present in table 2. Thirdly, we improved logging of the pipeline in the new version and updated an archived version of the repository. We also corrected the percent identity (PI) threshold from 0.5 to 50 to fit the definition of the percent in the text and we made corresponding changes in the code. Other minor changes include a change of a reference 1, a fix in a gene name Dot1l.
>
> **Any further responses from the reviewers can be found at the end of the article**

Introduction

Conventionally, microRNAs (miRNAs) are considered to suppress gene expression through RNA interference (RNAi) by binding complementarily to mRNAs, forming a RISC complex, and causing RNA degradation.[1] However, recent studies provide evidence that some miRNAs act in the opposite way – stimulating gene activation. Numerous studies have demonstrated the ability of miRNAs to up-regulate genes by targeting their promoters.[2,3] Ago1 from miRNA-Ago complex associates with the Ccnb1 promoter and miR-744 induces enrichment of RNA Pol II and H3K4me3 at the Ccnb1 transcription start site.[3] MiRNA let-7i interacts with the TATA-box of IL-2 gene and stimulates transcription initiation by contributing to the assembly of the pre-initiation complex.[4] Relatively fewer miRNAs demonstrated the ability to unconventionally target and activate enhancers, thus inducing genes regulated by these enhancers. MiR-24-1 acts as a modulator of the chromatin state of an enhancer. Furthermore, it increases p300 and RNA Pol II binding at the enhancer locus. The miR-24-1 actually originates from the enhancer locus. However, some genes regulated by other enhancers are also expressed at higher levels when miR-24-1 is transfected, and the enhancers of induced genes contain a sequence similar to the seed of the miRNA.[5] These observations suggest that other miRNAs might trigger enhancers and activate gene expression.

When miRNAs function as activators in a nucleus, different targeting mechanisms are possible: miRNA:DNA Watson-Crick duplex formation as well as miRNA:DNA Hoogsteen triple helix formation. Nuclear miRNA target prediction tools utilize are based on the idea that miRNA:DNA interaction requires an intact seed region.[6] MicroPIR2 predicts targets in mouse and human promoter regions.[7] Trident predicts miRNA:DNA Hoogsteen-type base pairings.[8] Some tools designed to predict conventional miRNA targets may also be utilized to find nuclear activational targets, e.g. miRanda.[9]

In this work, we report MIREyA (MIRnas functioning through Enhancer Activation), a pipeline for detection of miRNAs and their gene targets up-regulated through triggering their enhancer in the provided expression dataset. We applied MIREyA in order to identify and characterize activational miRNAs in *Mycobacterium tuberculosis* (Mtb) infected macrophage dataset. MiRNAs are important regulators in macrophage responses during Mtb infection that may act as host immunity agents as well as a tool exploited by pathogen agents to manipulate host cell pathways.[10] Yet, only the classical role of miRNA has been investigated so far in the context of Mtb infection.[11−19] Although multiple studies have shown the possibility of the activational role of miRNAs, this potential remains neglected in the study of miRNAs in bacterial infections. MIREyA found several miRNAs, which have not been shown to be functional in TB yet, suggesting it could be useful to find candidate activational miRNAs for further experimental validation. Mir155 has previously been shown to act as a negative regulator of essential mRNAs during TB,[20,21] but not as an activator.

Methods
Main steps of the algorithm

MIREyA aims to detect miRNAs with the potential to upregulate a gene via activation of its enhancer. It consists of three major steps:

1) The algorithm detects miRNA bound to the enhancers. This step is implemented with three different approaches described below.

2) The algorithm selects genes regulated by the enhancers selected in step 1). For this step the output from the first step is required as well as a table with enhancer:gene pairs where an enhancer is assumed to regulate the corresponding gene.

3) The algorithm calculates the Spearman's correlation coefficient (SCC) between the expression levels of miRNAs and genes regulated by corresponding enhancers selected in step 1), and estimates the p-value of the SCC with a Benjamini-Hochberg correction using the number of the miRNA:gene pairs for one miRNA (FDR < 0.05). The input data is gene expression data with sample size ≥ 8 and the output of the previous step.
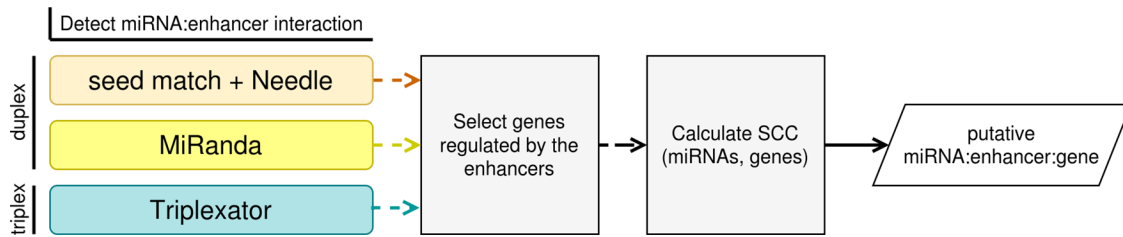
Figure 1. The outline of the implemented algorithm.

Figure 1 illustrates the workflow of our algorithm implemented in Python and R.

Prediction of miRNA-enhancer interaction

We speculate that in order to activate an enhancer, miRNA should bind to enhancer DNA. Since the mechanism of such binding is unclear, we decided to implement several reasonable prediction strategies. The first two strategies assume that miRNA binds to DNA forming an RNA:DNA double helix, while the third assumes RNA:DNA triple helix formation.

1) The first approach is to select an enhancer containing an exact match of the user-provided seed sequence of a miRNA, then expand each seed by 14 bp of the corresponding mature miRNA and align it to the enhancer with Needle tool[22] and keep only enhancers with the percent identity (PI) > 50 (PI defined as a percent of matches between miRNA and DNA region).

2) The second approach is to scan miRNA sequences against enhancer sequences and detect potential target sites with MiRanda.[9]

3) The third approach is to predict RNA:DNA triplexes between miRNAs and enhancers with Triplexator tool.[23] We relaxed the error-rate and lower-length-bound Triplexator default parameters in order to adjust the algorithm to work with extremely short miRNA sequences (error-rate=19, lower-length-bound=11).

The approaches are interchangeable, also the user can merge the results of all approaches to reflect multiple mechanisms of potential miRNA:DNA binding.

The main script to run the pipeline is src/run_mireya.py. Input and output files depend on choice of one of previously described approaches to predict miRNA-enhancer interaction and must be specified with the following flags:

-d/ --detection_mir_enh_interaction: approach to predict miRNA-enhancer interaction, accepts one of three possible values: seed_match_needle / miranda / triplexator

-e/ --enhancers: path to a fasta file with sequences of enhancers of interest

-o/ --output: full path to output directory

-ge/ --gene_expression: path to a .tsv file with normalized gene expression

-me/ --mirnas_expression: path to a .tsv file with expression of miRNAs of interest

-ei/ --enh_gene_interaction: path to a .tsv file with enhancers and corresponding genes they are assumed to regulate

-m/ --mature_mirnas: path to a fasta file with sequences of mature miRNAs of interest

Pipeline mode with the first approach to predict miRNA-enhancer interaction (-d seed_match_needle) requires additional input data which must be provided with flags:

-g/ --genome: path to a folder with fasta files with complete genome of the organism of interest: one chromosome per file

-s/ --seeds_mirnas_forward: path to a tab-delimited file with sequences of seeds of miRNAs of interest, without header with 2 columns: 1) general name of mirnas; 2) their seeds as DNA sequences (make sure that U are replaced with T)

-sr/ --seeds_mirnas_reverse_compl: path to a tab-delimited file with reverse complementary sequences of seeds of miRNAs of interest, without header with 2 columns: 1) general name of mirnas; 2) their seeds as DNA sequences (make sure that U are replaced with T)

-eb/ --enhancers_bed: path to bed file with coordinates of enhancers of interest

-ms/ --mature_mirnas_separate: path to a directory containing directories with mature miRNA sequences: one folder per miRNA containing fasta files with one mature sequence per file. Each directory with mature mirna fastas must be named exactly as in other input files. Names of fasta files will be used further as "mature_mirna" column in result tables.

MiRNAs must have the same names in -ge, -me, -ei, -s, -sr input data and for folders in -ms argument. Please, use example files to make sure the input is in correct format. MiRNA names for mature sequences must be the same in -m and -ms arguments.

Examples of command to run the main script, one for each approach to predict miRNA-enhancer interaction (names of files coinside with names of example input files in repository):

1. python src/run_mireya.py -d seed_match_needle -e data/enhancers.macrophages.Mtb.mm9.fasta -o out/seed_match_needle_out/ -ge data/DE_gene_expression.tsv -me data/DE_mirnas_expression.tsv -ei data/enh.gene.assoc.sign.tsv -m data/DE_mirna_mature_seqs.fa -g data/db -s data/seeds_seq_forward_short -sr data/seeds_seq_reverse_compl_short -eb data/enhancers.macrophages.Mtb.bed -ms data/mature_seqs/

2. python src/run_mireya.py -d miranda -e data/enhancers.macrophages.Mtb.mm9.fasta -o out/miranda_out/ -ge data/DE_gene_expression.tsv -me data/DE_mirnas_expression.tsv -ei data/enh.gene.assoc.sign.tsv -m data/DE_mirna_mature_seqs.fa

3. python src/run_mireya.py -d triplexator -e data/enhancers.macrophages.Mtb.mm9.fasta -o out/triplexator_out -ge data/DE_gene_expression.tsv -me data/DE_mirnas_expression.tsv -ei data/enh.gene.assoc.sign.tsv -m data/DE_mirna_mature_seqs.fa

Output file is called mir_enh_gene_trios.tsv. Example output file is placed in the out/directory of repository. One line in the output file corresponds to one trio of a mature miRNA, an enhancer and a gene ("mature_mirna", "Gene.Name", "enhancer" columns). Columns "corr (miRNA, gene)"and "p.value adj" correspond to Spearman's correlation coefficient of the miRNA and the gene expression and FDR. "PI" column corresponds to percent identity and is present only in the output of seed_match_needle approach.

Implementation
The pipeline is implemented as Python, R and bash scripts, and can be run with a master script run_mireya.py.

Operation
Python>=3.5 and r-base are expected to be pre-installed. Besides, two modes of the pipeline require the following tools installed: MiRanda, Triplexator. The pipeline was tested in Ubuntu and Ubuntu-based linux systems (Ubuntu>=16.04).

Use case
We applied MIREyA to three time-series (0, 4, 12, 24, 48, 96 hours) expression datasets (CAGE) of mouse bone marrow-derived macrophages infected with hypervirulent Beijing/W lineage *Mycobacterium tuberculosis* (Mtb) HN878 strain, 2-4 replicates per time point.[24] Each dataset corresponds to the time series after infection for macrophages of different phenotypes: not pre-stimulated (M0), interferon-γ stimulated (M1-polarized) and interleukin-4/interleukin-13 stimulated (M2-polarized). Only differentially expressed (DE) miRNAs and genes were considered. We obtained enhancer-gene interactome from[25] where an enhancer is predicted to regulate a gene if their expression levels correlate significantly and they belong to the same topologically associated domain (TAD).

We searched for candidate enhancers targeted by miRNAs with all three methods described previously and merged the results for further steps. CAGE enables one to estimate expression at the promoter level, while enhancers are associated with whole genes. As a proxy of gene expression, we used either an expression value of a promoter with the highest SCC (a) or summed up the expression values of all promoters of the gene (b). To reduce the number of false positive

predictions, we selected among miRNA:enhancer duplexes only such cases where (1) a duplex was predicted both by miRanda and Needle-based approach; (2) an enhancer was associated with several genes since in the original paper on miRNA-activated enhancers[5] one nuclear miRNA affected expression of multiple genes regulated by the triggered enhancer; (3) the miRNA-gene pair was obtained in both ways (a and b) to estimate expression of the gene. We also added miRNA-gene pairs with highly correlated expression levels (SCC ≥ 0.8) in any of four combinations of two methods to detect miRNA:enhancer interactions (seed match + Needle and miRanda) and two approaches to treat expression of different promoters (a and b). Among predicted miRNA:enhancer triplexes we selected cases where (1) an enhancer was associated with multiple genes; (2) both approaches to estimate gene expression (a and b) yielded this triplex.

Results

We applied MIREyA to three time-series datasets of Mtb-infected macrophages with 3 different phenotypes prior to infection. In M0 macrophages 10 miRNAs were differentially expressed (DE) in at least one time point compared to the state before infection (0h), in M1 there was no DE miRNA, while in M2 only Mir1956 was DE. Figure 2 and Table S1 (*Extended data*[43]) represent detected miRNA-enhancer-gene trios for M0 macrophages (*Extended data:* Table S2 for M2 macrophages[43]). We investigated the functions of the obtained genes and miRNAs and confirmed that many of them might be involved in the response to Mtb infection. We detected Mir155, which is vastly studied, and known to subvert autophagy in human dendritic cells[20] and to be a potential diagnostic marker of active tuberculosis.[21] Other miRNAs and their targets which we consider promising for further investigation are summed in Tables 1 and 2. Mir22, Mir221, Mir222 are annotated with high confidence.[26] Expression of only some promoters of Klf6 and BC016423 genes correlates with Mir22 expression, but we included them, because they have predicted triplexes with three different enhancers.
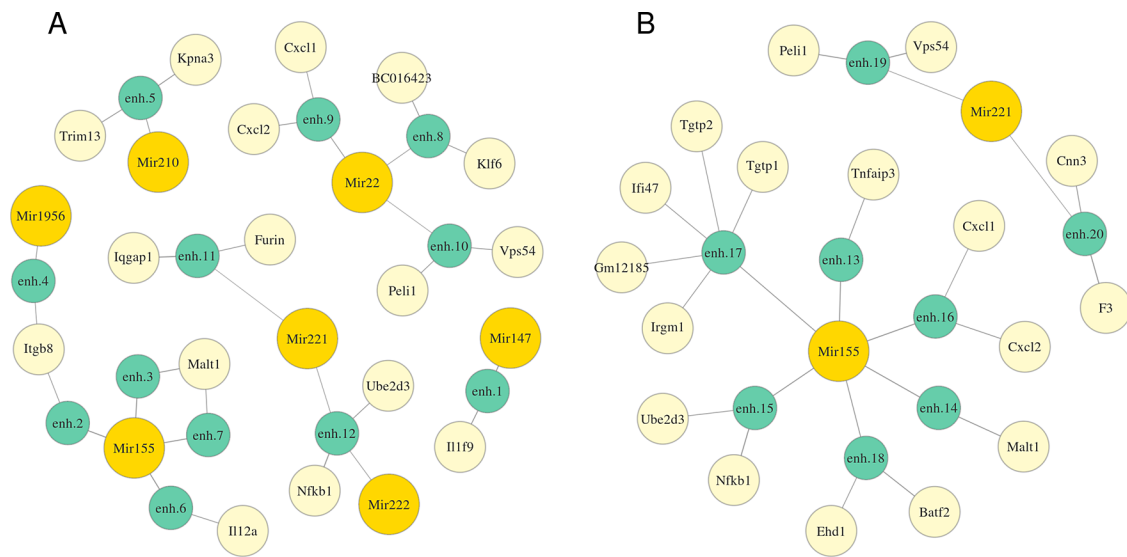

Figure 2. Predicted regulation graphs of miRNAs via RNA:DNA duplex (A) or RNA:DNA triplex formation (B) with an enhancer for M0 macrophages. The length of an edge between a miRNA and an enhancer or between an enhancer and a gene is in inverse proportion to the average correlation between the expression of the miRNA and the target gene or the enhancer and the gene respectively. Full coordinates of enhancers are available in Table S1.

Table 1. The most promising miRNAs predicted for M0 infected BMDMs to upregulate genes via enhancer activation.

miRNAs	genes	number of enhancers	mechanism of interaction of enhancer:miRNA
Mir22	Klf6, BC016423	3	duplex, triplex
Mir221, Mir222, Mir155	Nfkb1, Ube2d3	2	duplex, triplex
Mir22, Mir221	Peli1	2	duplex, triplex
Mir22, Mir221	Cxcl1, Cxcl2	2	duplex, triplex
Mir155	Malt1	2	duplex

Table 2. The most promising miRNAs predicted for M2 pre-stimulated infected BMDMs to upregulate genes via enhancer activation.

miRNAs	genes	number of enhancers	mechanism of interaction of enhancer:miRNA
Mir1956	Ccrl2	1	duplex
Mir1956	Dot1l	5	duplex
Mir1956	Cd14	1	duplex
Mir1956	Rab20	1	duplex
Mir1956	Ticam1	1	duplex
Mir1956	Tnfaip3 (A20)	2	duplex
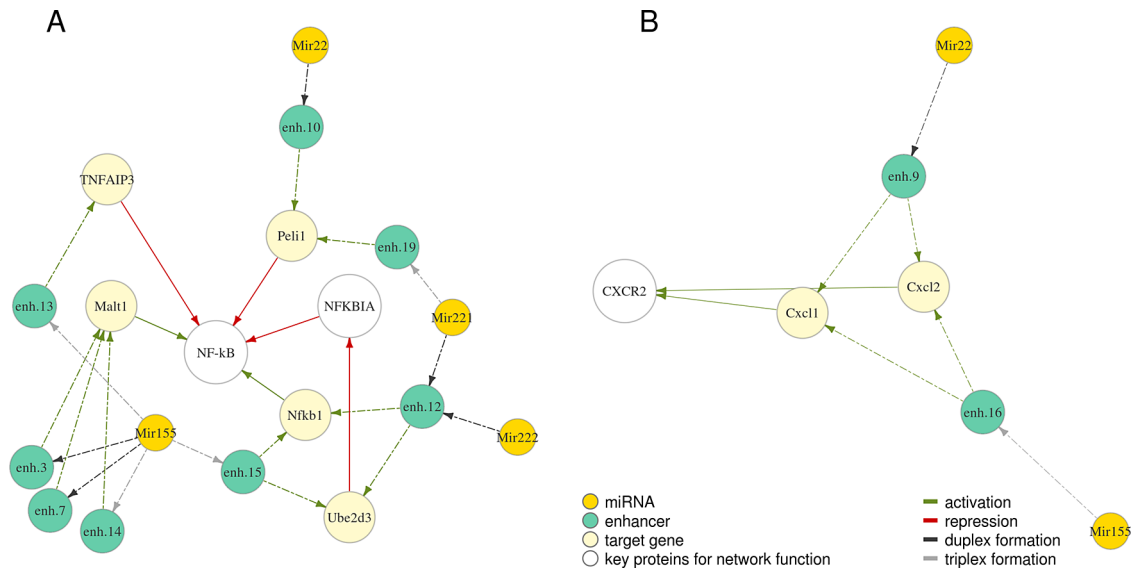
Figure 3. MiRNA-enhancer-gene trios detected for M0 macrophages include genes with related functions and form subnetworks A (related to NF-kB) and B (both genes encode chemokines which signal through CXCR2). Relationships are depicted with arrows of a different colour: green – activation, red – repression. Dashed arrows indicate predicted relationships, solid – known from published studies and described previously. Full coordinates of enhancers are available in Table S1.

We reconstructed regulatory networks for these miRNAs based on miRNA-enhancer-gene trios and investigated their potential role in the response to Mtb infection. Mir22-activated gene network is highly likely to be involved in Mtb response. Klf6, a potential target of Mir22, is a transcription factor essential for macrophage motility[27] and plays an important role in the regulation of macrophage polarization promoting M1 phenotype cooperatively with NF-κB.[28] Nfkb1 gene, a putative target of miRNAs Mir221, Mir222 and Mir155, encodes a subunit of NF-κB protein complex, a master transcription factor in macrophage immune responses. The human ortholog of Ube2d3, potentially regulated by the same miRNAs, facilitates polyubiquitination of NFKBIA (a member of the NF-kappa-B inhibitor family) stimulating its subsequent degradation.[29] A detected target of both Mir22 and Mir221, Peli1 regulates the NF-κB activity negatively and attenuates the induction of proinflammatory cytokines in T-cells.[30] Cxcl1 and Cxcl2, detected targets of both Mir22 and Mir221, are chemoattractants for neutrophils contributing to tissue inflammation.[31] Malt1, potentially up-regulated by Mir155, is known to activate NF-κB in lymphocytes.[32] Among targets of Mir1956 detected in M2 macrophages dataset we found the Ccrl2 gene encoding a chemokine receptor-like protein which is expressed at high levels in primary neutrophils and primary monocytes. Another Mir1956 target, Dot1l, is an H3K79 methyltransferase which facilitates the expression of IL-6 and IFN-β in macrophages.[33] Cd14 leads to NF-κB activation and inflammatory response,[34] Cd14 KO mice infected with Mtb are protected due to reduced inflammatory responses at the chronic stage.[35] Rab20 plays a role in the maturation and acidification of phagosomes and the fusion of phagosomes with lysosomes during mycobacterial infection.[36] Ticam1 is involved in native immunity against pathogens: it interacts specifically with toll-like receptor 3, activates NF-κB.[37] Tnfaip3 (A20) is an important regulatory protein that down-regulates NF-κB activity.[38]

We further investigated protein networks associated with the selected miRNAs. Proteins regulated by miRNAs Mir22, Mir155, Mir221, Mir222 are involved in regulation of NF-κB – a vital orchestrator of the response of the innate immune cells to pathogens[39] (Figure 3A). Cxcl1 and Cxcl2 regulated by Mir22 and Mir155 (Figure 3B) are chemokines which signal through CXC receptor 2 to attract neutrophils to the place of inflammation, which is essential to control tissue infection.[31]

Discussion

MIREyA aims to easily find candidate activating miRNAs which trigger an enhancer and genes up-regulated by the enhancer. New emerging methods for studying RNA–DNA interactome detect previously unknown miRNAs bound to chromatin[40,41] which might be promising for further experimental investigation in terms of understanding complex gene regulation networks in detail. Although high-throughput data on RNA:DNA interactions in several cell types but not macrophages are available now, we could not confirm MiRNAs detected with MIREyA for Mtb infection using either RADICL-seq[41] or GRID-seq,[42] which is unsurprising since the ncRNA:DNA interactions are highly cell-type-specific.[41]

Although the implemented algorithm considers multiple aspects of the suggested mechanism of gene activation, several factors remain unaccounted for. To run MIREyA a user requires *a priori* knowledge of enhancer-gene interactions. A common fast computational approach to determine gene-enhancer pairs is to use genes and enhancers co-localized in the genome, ignoring the long-distance spatial interactions. One of solutions is the approach suggested in:[25] to calculate the correlation between expression of genes and enhancers which belong to the same TAD.

Although the exact mechanism of RNA activation of enhancers remains unclear, we do know that miRNA as a mediator facilitates epigenetic modifications in the enhancer region. At this stage, MIREyA does not consider chromatin availability or any other epigenetic information. In order to reduce the number of false positive predictions, nuclear localisation of specific predicted miRNAs should be validated experimentally.

Despite the discussed limitations, using MIREyA we detected several promising miRNA candidates. We suggest that MIREyA provides a promising approach to select miRNAs which up-regulate genes by triggering enhancers for further experimental validation.

Conclusion

Our method extends the study of activational miRNAs and provides a basis for further research. The use case on Mtb-infected macrophages demonstrates the possibility of existence of novel miRNAs up-regulating gene expression.

Data availability

Underlying data

The CAGE time series expression datasets from the use case are available at https://fantom.gsc.riken.jp/data/ in mm9 Phase 2 release and can be selected to download with FANTOM5 Table Extraction Tool by «macrophage, TB infection» key words.

Extended data

Zenodo: MIREyA: Extended data, http://doi.org/10.5281/zenodo.4549445.[43]

This project contains the following extended data:

- Table S1: Detected miRNA-enhancer-gene trios for M0 macrophages

- Table S2: Detected miRNA-enhancer-gene trios for M2 macrophages

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Software availability

Pipeline available from: https://github.com/veania/MIREyA.

Archived pipeline as at time of publication: https://doi.org/10.5281/zenodo.5082988.[44]

License: Open Software License 3.0

References

1. O'Brien J, Hayder H, Zayed Y, et al.: **Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.** *Front Endocrinol.* 2018; **9**: 402.
   Publisher Full Text

2. Matsui M, Chu Y, Zhang H, et al.: **Promoter RNA links transcriptional regulation of inflammatory pathway genes.** *Nucleic Acids Res.* 2013; **41**: 10086–10109.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Huang V, Place RF, Portnoy V, et al.: **Upregulation of Cyclin B1 by miRNA and its implications in cancer.** *Nucleic Acids Res.* 2012; **40**: 1695–1707.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Zhang Y, Fan M, Zhang X, et al.: **Cellular microRNAs up-regulate transcription via interaction with promoter TATA-box motifs.** *RNA.* 2014; **20**: 1878–1889.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Xiao M, Li J, Li W, et al.: **MicroRNAs activate gene transcription epigenetically as an enhancer trigger.** *RNA Biol.* 2017; **14**: 1326–1334.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Liu H, Lei C, He Q, et al.: **Nuclear functions of mammalian MicroRNAs in gene regulation, immunity and cancer.** *Mol Cancer.* 2018; **17**: 64.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Piriyapongsa J, Bootchai C, Ngamphiw C, et al.: **microPIR2: a comprehensive database for human-mouse comparative study of microRNA-promoter interactions.** *Database.* 2014; **2014**: bau115.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Janowski BA, Huffman KE, Schwartz JC, et al.: **Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs.** *Nat Chem Biol.* 2005; **1**: 216–222.
   PubMed Abstract | **Publisher Full Text**

9. Betel D, Wilson M, Gabow A, et al.: **The microRNA.org resource: targets and expression.** *Nucleic Acids Res.* 2008; **36**: D149–D153.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Maudet C, Mano M, Eulalio A: **MicroRNAs in the interaction between host and bacterial pathogens.** *FEBS Lett.* 2014; **588**: 4140–4147.
    PubMed Abstract | **Publisher Full Text**

11. Kumar M, Sahu SK, Kumar R, et al.: **MicroRNA let-7 modulates the immune response to Mycobacterium tuberculosis infection via control of A20, an inhibitor of the NF-κB pathway.** *Cell Host Microbe.* 2015; **17**: 345–356.
    PubMed Abstract | **Publisher Full Text**

12. Wang J, Yang K, Zhou L, et al.: **MicroRNA-155 promotes autophagy to eliminate intracellular mycobacteria by targeting Rheb.** *PLoS Pathog.* 2013; **9**: e1003697.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Ma C, Li Y, Li M, et al.: **microRNA-124 negatively regulates TLR signaling in alveolar macrophages in response to mycobacterial infection.** *Mol Immunol.* 2014; **62**: 150–158.
    PubMed Abstract | **Publisher Full Text**

14. Ni B, Rajaram MVS, Lafuse WP, et al.: **Mycobacterium tuberculosis decreases human macrophage IFN-γ responsiveness through miR-132 and miR-26a.** *J Immunol.* 2014; **193**: 4537–4547.
    PubMed Abstract | **Publisher Full Text**

15. Ma F, Xu S, Liu X, et al.: **The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-γ.** *Nat Immunol.* 2011; **12**: 861–869.
    PubMed Abstract | **Publisher Full Text**

16. Dorhoi A, Iannaccone M, Farinacci M, et al.: **MicroRNA-223 controls susceptibility to tuberculosis by regulating lung neutrophil recruitment.** *J Clin Invest.* 2013; **123**: 4836–4848.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Kim JK, Lee H-M, Park K-S, et al.: **MIR144* inhibits antimicrobial responses against Mycobacterium tuberculosis in human monocytes and macrophages by targeting the autophagy protein DRAM2.** *Autophagy.* 2017; **13**: 423–441.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Guo L, Zhou L, Gao Q, et al.: **MicroRNA-144-3p inhibits autophagy activation and enhances Bacillus Calmette-Guérin infection by targeting ATG4a in RAW264.7 macrophage cells.** *PLoS One.* 2017; **12**: e0179772.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Gu X, Gao Y, Mu D-G, et al.: **MiR-23a-5p modulates mycobacterial survival and autophagy during mycobacterium tuberculosis infection through TLR2/MyD88/NF-κB pathway by targeting TLR2.** *Exp Cell Res.* 2017; **354**: 71–77.
    PubMed Abstract | **Free Full Text**

20. Etna MP, Sinigaglia A, Grassi A, et al.: **Mycobacterium tuberculosis-induced miR-155 subverts autophagy by targeting ATG3 in human dendritic cells.** *PLoS Pathog.* 2018; **14**: e1006790.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Wu J, Lu C, Diao N, et al.: **Analysis of microRNA expression profiling identifies miR-155 and miR-155* as potential diagnostic markers for active tuberculosis: a preliminary study.** *Hum Immunol.* 2012; **73**: 31–37.
    PubMed Abstract | **Publisher Full Text**

22. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet.* 2000; **16**: 276–277.
    PubMed Abstract | **Publisher Full Text**

23. Buske FA, Bauer DC, Mattick JS, et al.: **Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data.** *Genome Res.* 2012; **22**: 1372–1381.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Roy S, Schmeier S, Kaczkowski B, et al.: **Transcriptional landscape of Mycobacterium tuberculosis infection in macrophages.** *Sci Rep.* 2018; **8**: 6758.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

25. Denisenko E, Guler R, Mhlanga M, et al.: **Transcriptionally induced enhancers in the macrophage immune response to Mycobacterium tuberculosis infection.** *BMC Genomics.* 2019; **20**: 71.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep sequencing data.** *Nucleic Acids Res.* 2014; **42**: D68–D73.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Kim G-D, Das R, Goduni L, et al.: **Kruppel-like Factor 6 Promotes Macrophage-mediated Inflammation by Suppressing B Cell Leukemia/Lymphoma 6 Expression.** *J Biol Chem.* 2016; **291**: 21271–21282.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Date D, Das R, Narla G, et al.: **Kruppel-like transcription factor 6 regulates inflammatory macrophage polarization.** *J Biol Chem.* 2014; **289**: 10318–10329.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Gonen H, Bercovich B, Orian A, et al.: **Identification of the ubiquitin carrier proteins, E2s, involved in signal-induced conjugation and subsequent degradation of IkappaBalpha.** *J Biol Chem.* 1999; **274**: 14823–14830.
    PubMed Abstract | **Free Full Text**

30. Chang M, Jin W, Chang J-H, et al.: **The ubiquitin ligase Peli1 negatively regulates T cell activation and prevents autoimmunity.** *Nat Immunol.* 2011; **12**: 1002–1009.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. De Filippo K, Dudeck A, Hasenberg M, et al.: **Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation.** *Blood.* 2013; **121**: 4930–4937.
    PubMed Abstract | **Publisher Full Text**

32. Hailfinger S, Nogai H, Pelzer C, et al.: **Malt1-dependent RelB cleavage promotes canonical NF-kappaB activation in lymphocytes and lymphoma cell lines.** *Proc Natl Acad Sci U S A.* 2011; **108**: 14596–14601.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Chen X, Liu X, Zhang Y, et al.: **Methyltransferase Dot1l preferentially promotes innate IL-6 and IFN-β production by mediating H3K79me2/3 methylation in macrophages.** *Cell Mol Immunol.* 2020; **17**: 76–84.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

34. Haziot A, Ferrero E, Köntgen F, et al.: **Resistance to endotoxin shock and reduced dissemination of gram-negative bacteria in CD14-deficient mice.** *Immunity.* 1996; **4**: 407–414.
    PubMed Abstract | **Publisher Full Text**

35. Wieland CW, van der Windt GJW, Wiersinga WJ, et al.: **CD14 contributes to pulmonary inflammation and mortality during murine tuberculosis.** *Immunology.* 2008; **125**: 272–279.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Seto S, Tsujimura K, Koide Y: **Rab GTPases regulating phagosome maturation are differentially recruited to mycobacterial phagosomes.** *Traffic.* 2011; **12**: 407–420.
    PubMed Abstract | **Publisher Full Text**

37. Oshiumi H, Matsumoto M, Funami K, et al.: **TICAM-1, an adaptor molecule that participates in Toll-like receptor 3–mediated interferon-β induction.** *Nat Immunol.* 2003. pp. 161–167.
    PubMed Abstract | **Publisher Full Text**

38. Wertz IE, O'Rourke KM, Zhou H, et al.: **De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling.**

Nature. 2004. pp. 694–699.
PubMed Abstract | **Publisher Full Text**

39. Dorrington MG, Fraser IDC: **NF-κB Signaling in Macrophages: Dynamics, Crosstalk, and Signal Integration.** *Front Immunol.* 2019; **10**.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Gavrilov AA, Zharikova AA, Galitsyna AA, *et al.*: **Studying RNA–DNA interactome by Red-C identifies noncoding RNAs associated with various chromatin types and reveals transcription dynamics.** *Nucleic Acids Res.* 2020. pp. 6699–6714.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Bonetti A, Agostini F, Suzuki AM, *et al.*: **RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions.** *Nat Commun.* 2020;
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Li X, Zhou B, Chen L, *et al.*: **GRID-seq reveals the global RNA–chromatin interactome.** *Nat Biotechnol.* 2017. pp. 940–950.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

43. Elizarova A, Ozturk M, Guler R, *et al.*: **MIREyA: Extended data (Version 1.0.0) [Data set].** *Zenodo.* 2021.
Publisher Full Text

44. veania. (2021, July 8). veania/MIREyA: v1.0.2 (Version v1.0.2). Zenodo.
Publisher Full Text

F1000Research

Open Peer Review

Current Peer Review Status: ❓ ❓ ✔

Version 2

Reviewer Report 01 September 2021

https://doi.org/10.5256/f1000research.59225.r92860

✔ **Chaogang Shao**
College of Life Sciences, Huzhou University, Huzhou, China

My questions have been answered.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: miRNA data mining, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 05 July 2021

https://doi.org/10.5256/f1000research.31129.r87527

❓ **Chaogang Shao**
College of Life Sciences, Huzhou University, Huzhou, China

The authors have developed a pipeline, MIREyA, to predict miRNA gene-enhancer trios based on an expression dataset. This manuscript covers an important aspect of miRNA-target research by presenting a useful tool that technically sound. I have some questions:
1. A newly developed software needs to prove its function is correct before it is applied.

However, the manuscript lacks this important part. Though the authors applied the pipeline to primary murine macrophages infected by Mycobacterium tuberculosis and detected some miRNAs could up-regulate genes related to immune responses, these predicted results need to be proved by experiments. At present, we can't determine whether these regulatory relationships are correct, so the reliability of MIREyA can not be proved. The authors can use the published miRNAs with activation function, such as MiR-24-1, miR-744 and MiRNA let-7i, to verify the reliability of the pipeline.

2. Are there any miRNAs coming from primary murine macrophases affected by Mycobacterium tuberculosis that can bind to enhancers but reduce gene expression?

3. "keep only enhancers with the percent identity (PI) > 0.5 (PI defined as a percent of matches between miRNA and DNA region)". Will 0.5 be too low to increase false positives? What is the basis for the author to choose the parameter?

4. Are there any other sRNAs that share very similar sequence to miRNA and express even higher than miRNA in cells? These sRNAs may be involved in the regulation of enhancers too. Whether authors can improve the software to find these co-regulated sRNAs from high throughput sequence data, so as to build a more complete regulatory network.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: miRNA data mining, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Aug 2021

Anna Elizarova, Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, Russian Federation

Thank you for leaving the review.
As for the questions:

- A newly developed software needs to prove its function is correct before it is applied. However, the manuscript lacks this important part. Though the authors applied the pipeline to primary murine macrophages infected by Mycobacterium tuberculosis and detected some miRNAs could up-regulate genes related to immune responses, these predicted results need to be proved by experiments. At present, we can't determine whether these regulatory relationships are correct, so the reliability of MIREyA can not be proved. The authors can use the published miRNAs with activation function, such as MiR-24-1, miR-744 and MiRNA let-7i, to verify the reliability of the pipeline.

To find already studied miRNAs with this tool, we need an expression dataset for the same cell types and conditions as in the references. MiRNA expression is highly dependent on cell type and conditions, thus any other expression dataset would have a different set of expressed miRNAs which might not include miRNAs studied in the papers. Another limitation is that the expression dataset must include at least 8 samples so that correlation analysis would make sense. For now we are not aware of such a complete dataset.
In case of MiR-24-1, there was no expression dataset to run our pipeline, but the first step of the algorithm (miRNA/enhancer interaction prediction) would automatically find the enhancer from the article: the miRNA sequence is apparently present in the sequence of the enhancer in the reference 3 as the enhancer contains a gene sequence of the miRNA. We found an article (doi.org/10.1038/srep12987) reporting gene activation by miR-744 and an article (doi.org/10.1073/pnas.1803384115) reporting similar activation by miR-744 but the authors did not demonstrate that these miRNAs do that through enhancer activation. Lack of a known enhancer in a regulatory circuit makes the use of MIREyA impossible.

- Are there any miRNAs coming from primary murine macrophases affected by Mycobacterium tuberculosis that can bind to enhancers but reduce gene expression?

Since we focused on the activatory role of miRNA, we believe that the analysis of repressor miRNAs bound to enhancers is beyond the scope of the current paper.

- "keep only enhancers with the percent identity (PI) > 0.5 (PI defined as a percent of matches between miRNA and DNA region)". Will 0.5 be too low to increase false positives? What is the basis for the author to choose the parameter?

We believe that 0.5 is high enough. We chose the threshold by comparison with experimentally validated examples of miRNA interaction with RNA, where the minimum 11/22 miRNA bases represented an exact match. We hypothesized that such PI would be sufficient for miRNA interaction with DNA.

- Are there any other sRNAs that share very similar sequence to miRNA and express even higher than miRNA in cells? These sRNAs may be involved in the regulation of enhancers too. Whether authors can improve the software to find these co-regulated sRNAs from high throughput sequence data, so as to build a more complete regulatory network.

So far we have not seen any article describing sRNAs with similar to miRNA properties which

are able to regulate enhancers. We believe that for now this suggestion is beyond the scope of this work.

Competing Interests: No competing interests were disclosed.

Reviewer Report 29 June 2021

https://doi.org/10.5256/f1000research.31129.r87525

❓ **Xiangfu Zhong** 🆔

Karolinska Institutet, Stockholm, Sweden

Summary:
The manuscript describes a python pipeline to predict the interaction between miRNA, gene and enhancer, based on the expression data. Their interaction may help us have a better understanding of miRNA regulation mechanisms.

Comments and suggests for improvement:
1. Page 3, the methods section, the third step, "*the input data is gene expression data with sample size >=8*". Does this sample size requirement include replicates? Why does MIREyA require this number of sample size?

2. Page 5, in the "Use case" section, "*2-3 replicates per time point*". However, the reference 24 says "*Three-four biological replicates were prepared per time point*".

3. Can the authors provide some script or functions for users to process the raw data into MIREyA input format? For example, the first column in *data/DE_mirnas_expression.tsv*, is the miRNA names, but not in the format of mostly used.

4. The example expressions are normalized, it would be clear to mention when explaining *-ge* on page 4.

5. Improve the logging and exit message of running the pipeline.

Optional:
1. The current code looks like in the script format, wrapping them into package and release to pypi may improve the package popularity and availability. The code looks much like customer scripts for analysis for case use data, generation of them into wider usage.

2. Some genes have multiple copies across the genome, when calculating the correlation between gene expression and enhancer/miRNA, will copy/genomic coordinate be given to

the correlation test? At least 3 line expression data about Ccdc59 in the example expression data data/DE_gene_expression.tsv, which one is used in the correlation test?

3. I could not reproduce the first example 1 on page 5:

○ git clone git@github.com:veania/MIREyA.git

○ Rename the out folder within this repo code, then create another out for testing,

○ python src/run_mireya.py -d seed_match_needle -e data/enhancers.macrophages.Mtb.mm9.fasta -o out/seed_match_needle_out/ -ge data/DE_gene_expression.tsv -me data/DE_mirnas_expression.tsv -ei data/enh.gene.assoc.sign.tsv -m data/DE_mirna_mature_seqs.fa -g data/db -s data/seeds_seq_forward_short -sr data/seeds_seq_reverse_compl_short -eb data/enhancers.macrophages.Mtb.bed -ms data/mature_seqs/

First error in step 1:

○ Error: Unable to open file out/seed_match_needle_out//Mir146_seed_pos_in_genome. Exiting.

Even with this error, the pipeline keeps running without exit. For the step 2, I got another error:

○ File 'Mir146_enh_with_seeds.bed' has size 0. Returning a NULL data.table.

○ Execution halted

○ Calculation is finished. Please find the results in mir_enh_gene_trios.tsv

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: bioinformatics, microRNA annotation and target predication, epigenetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Author Response 02 Aug 2021

Anna Elizarova, Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, Russian Federation

Thank you for leaving the review.
As for the questions:

1. Page 3, the methods section, the third step, "*the input data is gene expression data with sample size* >=8". Does this sample size requirement include replicates? Why does MIREyA require this number of sample size?

Yes, the stated sample size includes replicates. We added this constraint because the pipeline relies on calculation of correlation coefficient. It is possible to try it on an expression dataset of smaller size but the result might be less reliable.

1. Page 5, in the "Use case" section, "*2-3 replicates per time point*". However, the reference 24 says "*Three-four biological replicates were prepared per time point*".

Authors of the original paper prepared up to 4 replicates, but sequenced only 2-4. To check the exact number of replicates sequenced for each time point one can browse FANTOM5 data (https://fantom.gsc.riken.jp/5/) or have a look at Supplementary table 1 in the reference 2 as well. We changed the description in our article to "2-4 replicates", since unstimulated macrophages' time point 24h has 4 sequenced replicates.

1. Can the authors provide some script or functions for users to process the raw data into MIREyA input format? For example, the first column in *data/DE_mirnas_expression.tsv*, is the miRNA names, but not in the format of mostly used.

It is unclear what exactly the reviewer means by mentioning the raw data. If the reviewer has in mind any specific format we can provide a script for the conversion. For now, one may use miRNA names in any format. The only thing to keep in mind is that they should be named consistently in all input files. Sequences of mature miRNAs may be downloaded from miRBase database: http://www.mirbase.org/ftp.shtml. Seed sequences of miRNAs could be retrieved from, e.g. http://www.targetscan.org/cgi-bin/targetscan/data_download.cgi?db=mmu_72.

1. The example expressions are normalized, it would be clear to mention when explaining *-ge* on page 4.

We specified that gene expression is expected to be normalized.

1. Improve the logging and exit message of running the pipeline.

Thank you for the suggestions. We added more logging in the latest release. The pipeline now exits if in seed_match_needle mode there is no match of miRNA seeds in the genome and genome as fasta files per chromosome are provided in incorrect format.

Optional:

1. The current code looks like in the script format, wrapping them into package and release to pypi may improve the package popularity and availability. The code looks much like customer scripts for analysis for case use data, generation of them into wider usage.

We believe that for now it is out of the scope of this work. This is an option to consider in

the future.

1. Some genes have multiple copies across the genome, when calculating the correlation between gene expression and enhancer/miRNA, will copy/genomic coordinate be given to the correlation test? At least 3 line expression data about Ccdc59 in the example expression data data/DE_gene_expression.tsv, which one is used in the correlation test?

The expression dataset we use represents promoters not genes obtained by CAGE-seq technique. It has been shown previously that on average each human gene has 5 promoters. So in our dataset "copies" represent different promoters of genes. The pipeline will consider them separately. If one has a similar dataset, they may choose a maximum correlation coefficient per gene after running the pipeline. Alternatively, it is possible to sum promoter expression into gene expression from the beginning and pass it to the pipeline using a flag -ge. In the example dataset we used promoter seperately.

1. I could not reproduce the first example 1 on page 5:

○ git clone git@github.com:veania/MIREyA.git

○ Rename the out folder within this repo code, then create another out for testing,

○ python src/run_mireya.py -d seed_match_needle -e data/enhancers.macrophages.Mtb.mm9.fasta -o out/seed_match_needle_out/ -ge data/DE_gene_expression.tsv -me data/DE_mirnas_expression.tsv -ei data/enh.gene.assoc.sign.tsv -m data/DE_mirna_mature_seqs.fa -g data/db -s data/seeds_seq_forward_short -sr data/seeds_seq_reverse_compl_short -eb data/enhancers.macrophages.Mtb.bed -ms data/mature_seqs/

First error in step 1:

○ Error: Unable to open file out/seed_match_needle_out//Mir146_seed_pos_in_genome. Exiting.

Even with this error, the pipeline keeps running without exit. For the step 2, I got another error:

○ File 'Mir146_enh_with_seeds.bed' has size 0. Returning a NULL data.table.

○ Execution halted

○ Calculation is finished. Please find the results in mir_enh_gene_trios.tsv

We repeated your actions on two computers and could not catch this error. To try to reproduce this error we need information about versions of your Linux distribution, Python, bedtools. Nevertheless, we did add a message about exiting for the case if this error occurs.

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 April 2021

Michiel Jan Laurens De Hoon 🆔

RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Summary:
This manuscript describes a novel software tool to detect interactions between miRNAs, enhancers, and the genes they regulate. Though the analysis performed by this software pipeline seems straightforward, it is a welcome addition as a bioinformatics tool, as the activation of enhancers by miRNAs has not been studied in much detail compared to the well-known mechanism of gene regulation by miRNAs via the RNAi pathway.

Comments and suggestions for improvement:
- Some of the references to the scientific literature are appropriate but quite old (for example, references 1, 9, and 26). Can these references be updated?

- On page 4, I assume that "percent identity (PI) > 0.5" should be "percent identity (PI) > 50".

- On page 5, the format and contents of the output file mir_enh_gene_trios.tsv is not clearly explained.

- In Table 2, should a column with the number of enhancers be added (same as in Table 1)?

Optional:
- In the results, what is the evidence that miRNA-gene interactions shown in Table 1 and 2 are true? If no such evidence is available, can the pipeline be used to rediscover the miRNA/enhancer/gene interactions described in reference 3, 4, and 5? I understand that this may be difficult as currently there are not many examples available together with suitable expression data sets, but perhaps it can be shown that at least the first step in the pipeline (detecting miRNA-enhancer interactions by Needle, MiRanda, or Triplexator) gives results consistent with what is described in these three references? That would make the paper more convincing.

- As a general point: Are both Python and R needed to run the pipeline? Could the pipeline be designed such that only Python or only R are needed?

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: I am a collaborator of Yulia A. Medvedeva in the FANTOM consortium, and we have been co-authors on several manuscripts in the past. I confirm that these competing interests have not affected my ability to write an objective and unbiased review of the article.

Reviewer Expertise: Computational genomics, bioinformatics, microRNA regulation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Aug 2021

Anna Elizarova, Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, Russian Federation

Thank you for leaving the review.
As for the questions:
- Some of the references to the scientific literature are appropriate but quite old (for example, references 1, 9, and 26). Can these references be updated?

We replaced reference 1 with a reference of a more recent review. The 9th reference is a reference to the original article of miRanda tool which is rather old but still functioning and suitable for our needs. The 26th reference is the most recent (2014) article to cite for miRBase database.
- On page 4, I assume that "percent identity (PI) > 0.5" should be "percent identity (PI) > 50".

We replaced 0.5 with 50 in the text of the article and we replaced PIs in the algorithm with values from 0 to 100 instead of 0 to 1.
- On page 5, the format and contents of the output file mir_enh_gene_trios.tsv is not clearly explained.

We added the following explanation to the main text of the article:
One line in the output file corresponds to one trio of a mature miRNA, an enhancer and a gene ("mature_mirna", "Gene.Name", "enhancer" columns). Columns "corr(miRNA, gene)" and "p.value adj" correspond to Spearman's correlation coefficient of the miRNA and the gene expression and FDR respectively. "PI" column corresponds to percent identity and is present only in the output of seed_match_needle approach.
- In Table 2, should a column with the number of enhancers be added (same as in

Table 1)?
We added a column with the number of enhancers to table 2.

Optional:

- ○ In the results, what is the evidence that miRNA-gene interactions shown in Table 1 and 2 are true? If no such evidence is available, can the pipeline be used to rediscover the miRNA/enhancer/gene interactions described in reference 3, 4, and 5? I understand that this may be difficult as currently there are not many examples available together with suitable expression data sets, but perhaps it can be shown that at least the first step in the pipeline (detecting miRNA-enhancer interactions by Needle, MiRanda, or Triplexator) gives results consistent with what is described in these three references? That would make the paper more convincing.

As for the evidence that miRNA-gene interactions shown in Table 1 and 2, we do understand that the output of our pipeline needs to be validated experimentally. The redescovering of the miRNA/enhancer/gene interactions described in reference 3, 4, and 5 is impossible since there's no expression dataset available in these references. MiRNA expression is highly dependent on cell type and conditions, thus any other expression dataset would have a different set of expressed miRNAs which might not include miRNAs studied in the papers. Repeating the first step (miRNA/enhancer interaction prediction) in the pipeline is impossible for the references 4, 5, since they describe other mechanisms of gene activation by miRNA without enhancer activation. The miRNA MiR-24-1 is apparently present in the sequence of the enhancer in the reference 3, because the enhancer contains a gene sequence of the miRNA as described in the article. Therefore an alignment of miRNA to the enhancer automatically will be successful.

- ○ As a general point: Are both Python and R needed to run the pipeline? Could the pipeline be designed such that only Python or only R are needed?

We started developing our pipeline as a set of R and bash scripts and added Python later due to the need of Biopython functions to work with fasta files. We will consider using only Python for MIREyA as a further step.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research