MIRROR 2.0: A PLATFORM FOR ASSESSING THE JOINT ACTION OF MICRORNAS IN CELL REGULATION

YITZHAK FRIEDMAN[*,†] and MICHAL LINIAL[*,¶]

[*]*Department of Biological Chemistry, Institute of Life Sciences,
The Hebrew University of Jerusalem,
Edmond J. Safra Campus, Givat Ram,
Jerusalem, 91904, Israel*

[†]*yitzhak.friedman@mail.huji.ac.il*
[¶]*michall@cc.huji.ac.il*

microRNAs (miRNAs) are short, non-coding RNAs that negatively regulate the levels of mRNA post-transcriptionally. Recent experiments revealed thousands of mRNA-miRNA pairs in which multiple miRNAs may bind the same transcript. These results raised the notion of miRNAs teamwork for a wide range of cellular context. miRror2.0 utilizes the miRNA-target predictions from over a dozen programs and resources and unifies them under a common statistical basis. The platform, called miRror2.0, considers the combinatorial regulation by miRNAs in different tissues, cell lines and under a broad range of conditions. A flexible setting permits the selecting of the preferred combination of miRNA-target prediction resources and varying the statistical parameters for the analysis. miRror2.0 covers six major model organisms including human and mouse. Importantly, the system is capable of analyzing hundreds of genes that were subjected to miRNAs' regulation. Activating miRror2.0 by introducing thousands of genes from miRNA overexpression experiments successfully identified the objective miRNAs. The output from miRror2.0 is a list of genes that is optimally regulated by a defined set of miRNAs. A symmetric application of miRror2.0 starts with a set of miRNAs, and the system then seeks the preferred set of genes that are regulated by that miRNA composition. The results from miRror2.0 are empowered by an iterative procedure called PSI- miRror. PSI-miRror tests the robustness of miRror2.0 prediction. It allows a refinement of the initial list of genes in view of the miRNAs that optimally regulate this list. We present miRror2.0 as a valuable resource for supporting cellular experimentalists that seek recovering the combinatorial regulation by miRNAs from noisy experimental data. miRror2.0 is available at http://www.mirrorsuite.cs.huji.ac.il.

Keywords: Non-coding RNA; RISC; microRNA-target pair; Prediction tool; Combinatorial regulation; StarBase; PAR-CLIP; miRBase.

[*,¶]Corresponding author.

1. Introduction

MicroRNAs (miRNAs) are short non-coding RNAs that negatively regulate gene expression post-transcriptionally. In humans, hundreds of miRNAs act in healthy and diseased tissues[1]. For example, miRNAs play a role as an additional layer of post-transcriptional regulation in stem cell differentiation, cell division, immunological cell function, stress response, organogenesis and cells' identity. Furthermore, miRNAs are involved in numerous human conditions (i.e., ageing, insulin resistance), pathogenesis (i.e., cancer, neurodegenerative diseases) and viral infection[2, 3]. At presence, mirBase is the most exhaustive collection of miRNAs with ~25,000 mature miRNA sequences from almost 200 organisms[4]. The over 2000 miRNAs known in human are estimated to target about half of the genes[5].

The basic mechanism for miRNA-target recognition is the base-pair complementarity at the RNA-induced silencing complex (RISC) complex[3]. While many experiments provided a cellular view for miRNA expression and regulation[6, 7], the understanding of the miRNA regulation in vivo is still fragmented[8]. Most experiments report on changes in the cell's transcriptome following the overexpression of a specific miRNA[9]. Other experiments monitor the expression of pre-selected targets by introducing the 3'-UTR segment on a reporter system[10]. The major effect of miRNAs is attributed to an enhanced degradation of the miRNA-RISC bound mRNAs[11]. Additionally, a detected effect on blocking translation from the miRNA-RISC bound mRNAs has been reported[12].

There are over a dozen miRNA-target prediction resources[13]. While all resources use the knowledge of a seed sequence complementarity, some algorithms also consider an imperfect hybridization, context-dependent features (e.g., accessibility of binding sites), evolution conservation, thermodynamic stability of the miRNA-mRNA duplex and more. Other algorithms use statistical models, including a machine-learning approach for assessing the efficacy of miRNA-target match[14]. Several assessment studies have demonstrated that the consistency among major miRNA-target predictors is rather poor, mostly due to the high number of false positives[15]. Analyzing the major miRNA-target databases (*e.g.,* DIANA-microT, TargetScan, Pictar and MirZ-ElMMO) resulted in a sensitivity of only 6-12%[16]. Based on this study, it was proposed that a combinatorial prediction performs worse than the prediction from one algorithm. TarBase and miRecords are manually compiled collections of miRNA-target pairs from in-vitro experiments[10] and the literature[17], respectively.

Direct measurements of miRNA bound to mRNA in a complex cellular system are provided by the recently developed technologies such as the high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) and the photoactivatable-ribonucleoside-enhanced crosslinking immunoprecipitation (PAR-CLIP) [18, 19]. The results from such experiments show that most of the genes were targeted by multiple miRNAs. Actually only 10% of the genes are regulated by a single miRNA[20]. The CLASH (crosslinking, ligation, and sequencing of hybrids) methodology detects the physical hybrid of miRNA and its target[21]. These experiments allow constructing an accurate miRNA-mRNA interaction map that takes into account teamwork by miRNAs.

Still, a substantial discrepancy remains when comparing complementary technologies[22]. Thus, an accurate model for miRNA regulation within cells is still missing.

In this research we expand the notion of the joint action by miRNAs, while benefiting from the ample amount of miRNA-target prediction resources. Along this principle, the MiRonTop[23], Diana-mirExTra[24] and GeneSet2miRNA[25] were developed. The basic algorithm was implemented in miRror1.0[26]. We recently showed that when using a statistical based argumentation for the miRNA-target pairing, a high precision for regulated genes is achieved without compromising on the specificity[27]. Herein, we demonstrate an advanced version of miRror that is designed to support experimentalists. Specifically, miRror2.0 suggests a coherent list of miRNAs that best explains an observed complex signature of down-regulated genes. In a dual view, an input of altered miRNAs in a cell leads to a list of genes that are the most likely targets of such miRNAs. The system supports the analysis of miRNA-target paring for major model organisms.

2. Methods

2.1. *miRror outline*

miRror2.0 is a software and a platform by which an experimentalist can gain biological insights on sets of molecular entities. miRror2.0 provides the main tool while PSI-miRror (Probability Supported Iterative miRror) operates to improve and refine the prediction results from miRror2.0. The input for miRror2.0 and PSI-miRror are (i) sets of miRNAs and (ii) gene-targets / proteins. Such sets are often a product of large-scale experiments such as miRNA profiling, mass spectrometry or gene expression screenings. The rich set of optional parameters is mostly an addition to the original version of miRror1.0[26]. miRror2.0 allows expanding the analysis for tissues, cell lines, level of expression and predictors' scores and the concept of miRNA families. The results are ranked according to the internal score (see section on miRIS).

2.2. *miRNA-target databases*

The miRNA-target prediction databases (coined MDBs) that are included in miRror are: TargetRank database[28]; microCosm that is based on miRanda algorithm[29]; PicTar[30]; DIANA–MicroT[31]; PITA[32]; EIMMO-MirZ[33]; microRNA.org[34]; miRDB[35]; TargetScan[9] and somewhat less extensive resources of RNA22 (cbcsrv.watson.ibm.com/rna22.html) miRNAMap2[36] and the meta-predictor MAMI (http://mami.med.harvard.edu). PITA, TargetScan, TargetRank and PicTar provide separated lists of predictions with a different level of coverage and specificity. A total of 15 MDBs are associated with humans. MirZ[33], microRNA.org[34] and miRBase[37] also provide information related to miRNA expression profiles for a large number of tissues and cell lines. Each of the MDB is characterized by the specific pairing between the miRNAs and targets (Fig. 1A). For example, TargetScan covers ~1500 miRNAs and ~25,000 genes, while miRNAMap2 covers only 470 miRNAs and ~6000 genes. mirRror2.0 supports *D. melanogaster*, *H.*

sapiens, M. musculus, R. norvegicus, C. elegans and *D. rerio* (marked as D.m, H.s, M.m, R.n, C.e, and D.r in Fig. 1A).

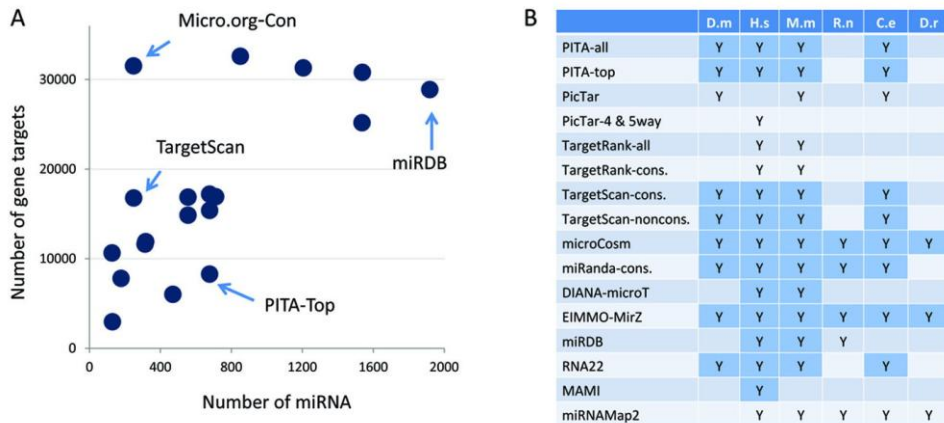|	D.m	H.s	M.m	R.n	C.e	D.r
PITA-all	Y	Y	Y		Y	
PITA-top	Y	Y	Y		Y	
PicTar	Y		Y		Y	
PicTar-4 & 5way		Y				
TargetRank-all		Y	Y			
TargetRank-cons.		Y	Y			
TargetScan-cons.	Y	Y	Y		Y	
TargetScan-noncons.	Y	Y	Y		Y	
microCosm	Y	Y	Y	Y	Y	Y
miRanda-cons.	Y	Y	Y	Y	Y	
DIANA-microT		Y	Y			
EIMMO-MirZ	Y	Y	Y	Y	Y	Y
miRDB		Y	Y	Y		
RNA22	Y	Y	Y		Y	
MAMI		Y				
miRNAMap2		Y	Y	Y	Y	Y

Fig. 1. miRNAs and target-genes databases supported by miRror2.0. (A) A total of 17 pre-compiled MDBs. Selected MDBs are marked. (B) The MDBs marked according to organisms. 'Y' marks the MDBs that provide information for the indicated organism. Gray background indicates than an internal scoring ranking method exists for the indicated MDB.

2.3. *Nomenclature and ID mapping*

The capacity of miRror2.0 to combine MDBs and resources is based on converting the miRNAs and gene-targets identifiers. miRror uses the RefSeq identifier as a central entry. miRror supports UniProtKB accession and IDs, Official gene symbol, Entrez ID, Flybase accession and Ensembl ID. Conversion was performed using BioMart (www.ensembl.org/biomart), ID converter (idconverter.bioinfo.cnio.es) and UniProtKB ID mapping (http://www.uniprot.org/mapping). A total of 15 identifiers are supported for the identifiers' exchange among MDBs. The match of RefSeq to the original accession of the MDBs reaches 98% for human and mouse and is lower for the other organisms. Currently, gene isoforms are not explicitly supported.

2.4. *miRror principles and concept*

The miRror2.0 platform is used to connect a gene list to the minimal set of preferred miRNAs sets or a miRNA collection to a set of genes that are most likely to be regulated by this set. The core of the statistical basis of miRror is the miRtegrate algorithm (Fig. 2). In a nutshell, miRtegrate calculates the probability of matches between the user gene-list as compared to all the genes that are selected by the relevant miRNA-targets predicting MDBs (Fig. 1A). The probability of the miRNAs' interaction with the input gene set yields a calculated p-value threshold for all probabilities that are more significant than the indicated threshold. Formally, the probability of the miRNA interaction with the input

gene set as opposed to the rest of the genes in that MDB is calculated. We applied a statistical threshold to ensure that the contribution of any MDB that covers only a small number of miRNAs with high specificity remains significant. In the case of MDB that over-predicts (*i.e.,* each miRNA matches almost any of the genes), our statistical threshold ensures that such MDB will have a reduced impact on the results. Calculating the p-value for genes as input was performed according to the hypergeometric distribution. N is the total number of genes in the database, n is the number of genes predicted for the specific miRNA, m is the number of genes in the input set and k is the number of genes that appear in both the input set and the predicted miRNA targets, the probability of such k is:

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \tag{1}$$

P-value is obtained by summing overall $X \geq k$. A correction for multiple testing was included using the FDR procedure. The p-value threshold that was used in this study is <0.05 (this remains a free parameter, it can be changed by the user).

2.5. *miRror2.0 input*

The input for miRror is a set of genes or a set of mature miRNAs. The user must select one of the 6 supported organisms (human, mouse, rat, worm, fly and zebrafish). The relevant MDBs for each of the organisms and the availability of a ranking and scoring method by the individual MDB are marked in Fig. 1B.

The user can activate miRror2.0 in the miR2Gene or Gene2miR operational modes. Fig. 2 illustrates the workflow for miRror2.0 application in the Gene2miR mode while emphasizing the optional selection of filters and parameters. In the Gene2miR mode, a set of genes is loaded as the input. A set of miRNAs that is more significant than the pre-selected p-value will be reported.

The same logic is applied in the miR2Gene mode. Following a selection of an organism and the operational mode, several optional choices are available:

(1) The tissue of interest / preferred cell-line. These collections are based on the atlas of gene expressions from of all mouse and human tissues[38]. Tissues are annotated by the gene expression data from FlyAtlas[39] and the atlas of gene expressions[40]. The probability of the miRNAs' interaction with the input gene set is recalculates for the selection of cell-line or tissue. Information concerning the selection of cell-lines or tissues is missing for the miR2Gene mode.

(2) All genes / only highly expressed (above a predetermined threshold). Data extracted from the atlas of gene expressions[40]. Typically this filter reduces the original list of genes by ~30%.

(3) The top scoring miRNAs according to each MDB internal scoring system. Note that some MDBs do not support an internal scoring (Fig. 1B).

(4) The user should select the preferred MDBs (15 in the case of human and 3 in the case of zebrafish). While any combination of MDBs is accepted (according to Fig. 1), a recommended for a minimal set is suggested as default. Only MDBs that comply with the user selection (*e.g.,* organism, top scoring predictions) will be presented.

Before activation a miRror2.0 search, the user may change several of the default parameters:
- The p-value threshold (default value is set to p-value <0.05).
- The minimal number of supporting MDBs (default, 2 MDBs)
- The number of input hits (*i.e.,* the minimal number of hits from the query set that are supported, default, 2 hits from the query list).
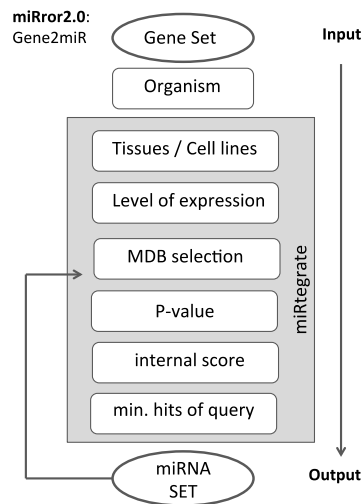

Fig. 2. A scheme of miRror2.0 operating at the Gene2miR mode. miRtegrate is the core of miRror2.0. It provides the statistical basis for the analysis. The parameters and filters can be changed for reanalysis option.

By changing these free parameters, the user can activate a relaxed or a strict search protocol. For example, the user can increase the minimal agreement from at least 2 MDBs to 3 MDBs. We encourage the user to control the size of the input by repeating the miRror2.0 protocol with a varying fraction of regulated genes (Fig. 2, left arrow). A balance between an increased size of the output predictions and high specificity in the predictions is achieved by using the optional parameters. Recall that the user can control the stringency of the analysis by activating a 'reanalyze' option on the same query.

2.6. *miRror internal score (miRIS)*

The miRIS is an internal score that balances between the consistency of prediction among the selected MDBs and the fraction of hits from the input. miRIS values range from 0 to 1

with an equal contribution of these two components to the score. The fit between the number of MDBs that support the prediction and the ration of the input hits from the input size is moderate (R^2=0.305). Thus the combination of these individual measures boosts the significance of the score. The benefit of ranking the results by miRIS was confirmed[27].

$$miRIS = \frac{\frac{DB\ hits}{\#DB} + \frac{query\ hits}{query\ size}}{2} \tag{2}$$

Due to the definition of a minimal number of MDBs that support the prediction (default=2), miRIS is always >0. For example, in a search for 10 MDBs the minimal miRIS value is 0.1 (for hits that are supported by only 2 MDBs).

A fit for the p-value of a prediction (determined as the minimal p-value that was reported from any of the MDBs that support such prediction, transformed to 1-log scale) and the number of MDBs is significant (R^2=0.496). A similar fit was calculated also for the p-value and the input hits (R^2=0.419). While the p-value is a significant quantitative estimate for each prediction, we had not included it in the miRIS score. The user can use the p-value to filter the results.

2.7. *Web implementation*

The web design scheme of miRror2.0 is based on open source software. Briefly, miRror is programmed in Perl and the web server is hosted on an Apache 2.0 HTTP server. The XML files were implemented in PHP language scripts that provide results upon request once a job is finished. miRror2.0 web application is implemented in an operating-system independent way and has been successfully tested in all major browsers. The system will be updated once a year and a number of recently developed predictor resources will be added. Additionally, the miRIS score will include a parameter to allow the user to rebalance between components of the score (*i.e.,* the number of supporting MDBs and the number of hits from the query).

3. Results and Discussion

3.1. *miRror2.0 output*

The modes of operation of miRror2.0 are named according to the nature of the input and the output. For example Gene2miR stands for a set of Genes (input) that leads to a set of miRNAs (output). The goal of these operational modes is to find the most likely set of molecules that are regulated by the input composition. All the modes apply with the same logic. miRror2.0 operates in 2 modes (Gene2miR and miR2Gene). Fig. 3 illustrates the results from the Gene2miR operational mode:
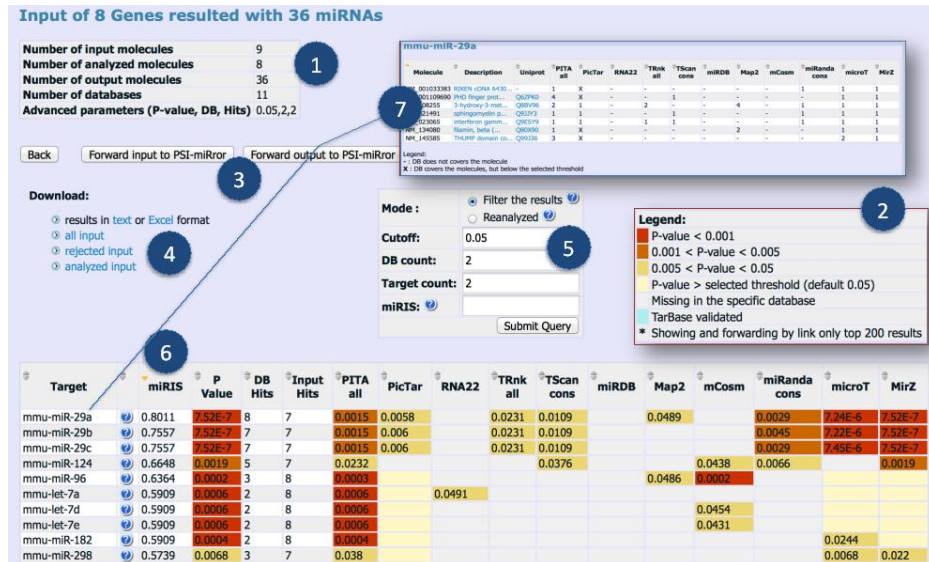
Fig. 3. miRror2.0 output. The main page of the results at the Gene2miR operational mode is shown. (1) A summary table. (2) A legend for the p-value thresholds. (3) Forwarding option to PSI-miRror. (4) Download options. (5) Restricting the output by changing the parameters. (6) The main result table, sorted by miRIS. (7) A zoomed table for each one of the predictions. For details see 'help' in www.mirrosuite.huji.ac.il. In this example, there are 11 MDBs (1) that were activated. The top ranked prediction is mmu-miR-29a that is supported by only 8/11 MDBs (marked as DB hits) and only 7/8 are marked as Input hits. Thus, the calculated miRIS is: ((7/8)+(8/11))/2=0.8011.

(1) A table that summarizes the user selections and the parameters that were applied for the search (Fig. 3, marked 1). In addition, the actual number of analyzed molecules is reported. Note that the number of the analyzed molecules may be somewhat smaller from the original input due to a mapping failure.

(2) A main table of results. In the case of Gene2miR mode, this table includes the miRNAs that are associated with the number of input hits and the number of MDBs that support the pairing of the genes in the input, and the miRNAs output list. The table is colored according to the p-value that is derived from the minimal p-value from any of the selected MDBs (Fig. 3, marked 2). Recall that the molecules reported in the result table are always more significant than the preselected p-value threshold. In the main table we present miRIS (Fig. 3, marked 6). The default sorting is by miRIS. The user can resort the table by any of its columns. An additional feature to the output table is the miRNA-target pairs that are reported as validated predictions by the TarBase resource[10].

(3) A zoomed table. For each of the resulted miRNA, a table is presented to indicate the actual MDBs that support that miRNA (Fig. 3, marked 7). The actual number of binding sites that are reported by each of the supported MDBs is shown. The user can

forward the miRror2.0 results to PSI-miRror for further analysis (Fig. 3, marked 3). Recall that the number of binding sites is not included in calculation of the miRIS.

We provide rich options for downloading and forwarding the results for additional analysis (Fig. 3, marked 4).

The user can activate number of filters for presenting a selected set of the full results. The option of "Reanalysis" (Fig. 3, marked 5) is used to increase or decrease the stringency of the initial analysis. Evidently, a selected of a more significant statistical threshold for the reanalysis applies that each of the MDB must meet this stringent threshold. For example, setting a reanalysis with a selected p-value of 0.001 has the potential to drastically reduce the number of the original predictions (default p-value <0.05).

3.2.   *Test case 1: Mildly down-regulated gene set*

The resource of miRror2.0 integrates most miRNA-targets predictors while assuming a combinatorial action of miRNAs toward a set of targets. We tested whether the miRror approach can identify the dominating role of the appropriate miRNA using gene expression data alone. The miRror algorithm relies on a statistical unification for a dozen miRNA-target predictors.

We illustrate the task of activating miRror2.0 using the default parameters. We used the list from an overexpression experiment for hsa-miR-124 (GSE6207, HepG2 cells, 24 hours post transfection)[41]. We collected all down-regulated genes at a permissive repression level of 1.2 fold relative to control cells. The 100 most significant down-regulated genes were used as the input. The miRIS indicated hsa-miR-124 and hsa-miR-506 as the two top predictions. Interestingly, these two miRNAs belong to the same family according to TargetScan but are not reported by the other used MDBs. This case test shows the potential of miRror to recover the miRNA regulation from a rich, unbiased set of down-regulated genes.

3.3.   *Test case 2: Thousands of down-regulated genes as input*

In some instances, the signature of down-regulated genes is reflected by thousands of genes that were altered by overexpressing a single miRNA. For hsa-miR-1 (GSE22002, Hela Cells) more than 3200 genes were repressed to a modest level of 1.2 fold (Table 1). We activated miRror2.0 with the minimal list of only 1% of the repressed genes (32 genes) and gradually included more genes to the input. We show that even with >3200 genes, the prediction output marked hsa-miR-1 among the top ranking miRror2.0 predictions (Table 1).

Table 1 presents such analysis for hsa-miR-124, hsa-miR-1 and hsa-miR-155 (data retrieved from GEO series, GSE2002). The down-regulated genes were partitioned to inclusive groups (top 1%, 2.5%, 5%, 10%, 25%, 50%, 75% and all). The high coverage of genes from the query is a reliable measure for the successful predictions as implemented by miRIS. Identification of the overexpressed miRNA from the profile of

the repressed genes turned out to be surprisingly robust. Recovering the subjected miRNA was achieved even by introducing hundreds of genes (Table 1). Activating miRror2.0, using a similar size random sets, failed to provide appropriate predictions (not shown). For has-miR-155, a quality prediction is limited to a query that includes only 25% of all repressed genes (593 genes). Including a larger fraction of the down-regulated genes added noise and hsa-miR-155 could not be recovered among the few top predictions.

Table 1. miRror2.0 rank of overexpressed miRNAs

	% of down regulated genes	# of hits from query	Total genes in query	Rank in top 200 miRNAs
hsa-miR-1	1	30	32	2
	2.5	73	82	2
	5	131	164	1
	10	241	329	1
	25	488	823	2
	50	884	1647	2
	75	1202	2471	3
	100	1505	3295	3
hsa-miR-155	1	15	23	1
	2.5	41	59	1
	5	78	118	2
	10	152	237	2
	25	333	593	3
	50	569	1187	11
	75	757	1781	11
	100	939	2375	37
hsa-miR-124	1	18	24	1
	2.5	43	60	1
	5	77	120	1
	10	154	240	2
	25	359	601	2
	50	665	1203	2
	75	942	1805	3
	100	1197	2407	4

3.4. *The PSI-miRror*

PSI-miRror is an additional capability of the miRror2.0 that tests the robustness of the results and allows an iterative refinement of the output. PSI-miRror includes for each iteration a consideration of "many to many" in addition to "many to one" that is dominating by miRror2.0. This is achieved by increasing an internal PSI-miRror score that favors the miRNAs that have a high match with the set of the input genes.

The PSI-miRror input is identical to the input of miRror2.0. Each complete iteration (at the Gene2Gene and miR2miR modes) starts with a set of molecules (genes or miRNAs, respectively) and ends with a refined list of the same type of molecules as the input. For example, a full cycle of iteration in the Gene2Gene mode implicates that an input of a gene set is used to predict a set of miRNAs that is then used as input for miRror2.0 that predicts a set of the most likely genes. For simplicity, one can see this iteration as a continued run of miRror2.0 in a Gene2miR followed by a miR2Gene mode. The Gene2miR and miR2Gene modes are based on activating an incomplete iterative cycle.

Fig. 4 illustrates Gene2Gene mode with 2 iterations. The results of the iterations are presented as a Venn diagram that shows the difference between the original set (g1) and the final gene set (g3). The set of genes that was stable throughout the PSI-miRror operation is shown as an overlap in the Venn diagram. The system provides also the lists of genes that were removed / added by the iterations.

In addition to the parameters that are used by miRror2.0, the user can select the parameters for the iterations:

• The number of iterations (the default is set for two iterations).
• A relaxation level for each of the iterations (e.g., defining the maximal fraction of the molecules that can be added at each iteration). The lower is this fraction the more conservative is the predictions.
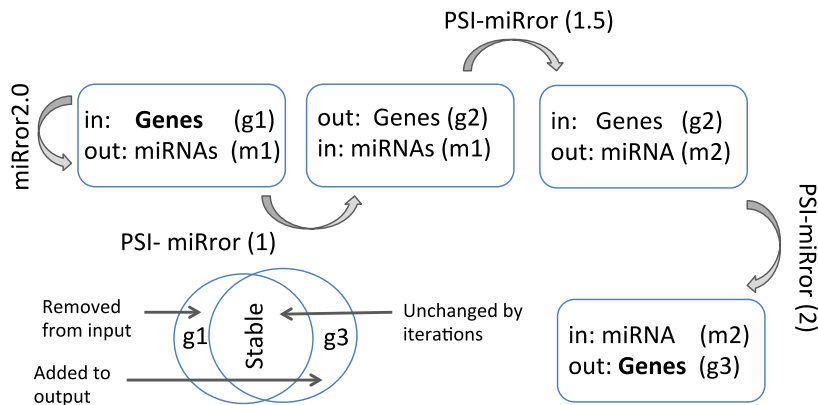• A complete (e.g., miR2miR) or a partial operational mode (e.g., miR2Gene).

Fig. 4. PSI-miRror outline. The PSI-miRror is an iterative application that is activated in two modes: miR2miR and the miR2Gene. The g3 is the output of the g1 input following 2 complete iterations. In the Gene2miR mode, the Gene set input (g1) is subjected to incomplete iteration cycles (*e.g.,* the output is a miRNAs set, marked m2). Removal or addition of entities from the original input sets is based on an internal score of the PSI-miRror algorithm.

3.5. *Test case 3: Refining the miRror2.0 predictions by PSI-miRror*

A numerical toy example for a Gene2Gene mode of PSI-miRror is described. A set of 20 genes (g1, Fig. 4) is used as an input. The results from activating the miRror2.0 protocol are 10 miRNAs (m1). In the first iteration, these 10 miRNAs are used as an input. miRror results report on 18 Genes (g2), only 15 of these genes are included in the original set (intermediate results are not reported by PSI-miRror). In the second iteration, the 18 genes are used as an input set and the results are 12 miRNAs (m2). In the same iteration, these 12 miRNAs set are again used as input in PSI-miRror and the result indicated 22 genes (g3), with only 19 of them that are included in the original set. The

results are illustrated as a Venn diagram. The different sections show the overlap of 19/20 genes (g1=g3), additional genes that were not included in the original gene set and genes that were removed from the original list are listed. When PSI-miRror was activated on complex biological data, the resulted overlapped molecules (miRNA or genes) that following the iterative cycles are most likely to represent the most stable set for further analyses.

The user may decide to retrieve the results without completing the last phase of the iterative cycle. In such scenario, an input of a gene set (g1) will provide a list of miRNAs (m2). This procedure can be operated in a symmetrical fashion where the input is miRNAs' set and the output is a refined set of genes.

Numerous benefits are associated with PSI-miRror:

(1) It can recover a gene that has similar characteristics in view of its miRNAs regulation, yet it was not included in the original set.
(2) miRNAs that belong to the same family that were missing from the experimental setting are often recovered.
(3) It allows a systematic testing of the coherence of an input set.
(4) It finds a simple explanation for miRNA regulation by removing the molecules that are inconsistent with such explanation

3.6. *An interpretation for miRror2.0 prediction*

miRror2.0 provides options for downloading and forwarding the results for a downstream analysis. The gene lists may be forwarded to external resources for assessment the functional annotations and a potential pathway. The tools that are selected for downstream analysis are DAVID[42], PANDORA[43], STRING[44] and the Reactome[45]. Activating any of these tools allows a deeper investigation of the predicted genes in term of the protein-protein and functional interaction (STRING), the regulatory pathways and processes (Reactome and DAVID), and the functions, domains and structural annotations (DAVID and PANDORA). The downloading options allow forwarding the result to any external analysis tool.

3.7. *The miRror2.0 outlook*

miRror2.0 is a platform that empowers the experimental biologists in gaining insights from a broad range of experimental protocols. Examples are the differentially expressed miRNAs, from profiling experiment, protein lists from mass spectrometry experiments, and genes from expression arrays. miRror2.0 is based on a "many-to-one" approach in which a set of miRNAs as input is optimized for the minimal set of targets that are maximally regulated by this set. Similarly, this applies for a set of genes as input. The "many-to-many" optimizations are performed by the PSI-miRror approach that yields a refined set of the input molecules by several iterative cycles. Activating the iterative cycles can lead to more accurate predictions. miRror2.0 website integrates many of the high quality miRNAs-target predictors. The numerous optional filters and parameters

allow the user to freely activate a strict or a relaxed protocol, and consequently control the extent of the predicted results. Differently from some meta-servers, miRror provides an integrative, statistically based platform and presents miRIS as a combined score for the prediction of miRNA combinatorial regulations

4. Acknowledgements

5. References

1. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414 (2007).
2. Alvarez-Garcia, I. & Miska, E.A. MicroRNA functions in animal development and human disease. *Development* **132**, 4653-4662 (2005).
3. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).
4. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**, D152-157 (2010).
5. Bentwich, I. Prediction and validation of microRNAs and their targets. *FEBS Lett* **579**, 5904-5910 (2005).
6. Thomson, D.W., Bracken, C.P. & Goodall, G.J. Experimental strategies for microRNA target identification. *Nucleic Acids Res* **39**, 6845-6853 (2011).
7. Guo, L. et al. Consistent isomiR expression patterns and 3' addition events in miRNA gene clusters and families implicate functional and evolutionary relationships. *Mol Biol Rep* **39**, 6699-6706 (2012).
8. Brodersen, P. & Voinnet, O. Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol* **10**, 141-148 (2009).
9. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).
10. Vergoulis, T. et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* **40**, D222-229 (2012).
11. Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).
12. Filipowicz, W., Bhattacharyya, S.N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102-114 (2008).
13. Min, H. & Yoon, S. Got target? Computational methods for microRNA target prediction and their extension. *Experimental & molecular medicine* **42**, 233-244 (2010).
14. Saito, T. & Saetrom, P. MicroRNAs--targeting and target prediction. *New biotechnology* **27**, 243-249 (2010).
15. Marin, R.M. & Vanicek, J. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res* (2010).
16. Alexiou, P., Maragkakis, M., Papadopoulos, G.L., Reczko, M. & Hatzigeorgiou, A.G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25**, 3049-3055 (2009).
17. Xiao, F. et al. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**, D105-110 (2009).

18. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479-486 (2009).
19. Hafner, M., Lianoglou, S., Tuschl, T. & Betel, D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **58**, 94-105 (2012).
20. Yang, J.H. et al. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res* **39**, D202-209 (2011).
21. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654-665 (2013).
22. Wen, J., Parker, B.J., Jacobsen, A. & Krogh, A. MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action. *RNA* **17**, 820-834 (2011).
23. Le Brigand, K., Robbe-Sermesant, K., Mari, B. & Barbry, P. MiRonTop: mining microRNAs targets across large scale gene expression studies. *Bioinformatics* **26**, 3131-3132 (2010).
24. Alexiou, P. et al. The DIANA-mirExTra web server: from gene expression data to microRNA function. *PLoS One* **5**, e9171 (2010).
25. Antonov, A.V., Dietmann, S., Wong, P., Lutter, D. & Mewes, H.W. GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res* **37**, W323-328 (2009).
26. Friedman, Y., Naamati, G. & Linial, M. MiRror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics* **26**, 1920-1921 (2010).
27. Balaga, O., Friedman, Y. & Linial, M. Toward a combinatorial nature of microRNA regulation in human cells. *Nucleic Acids Res* **40**, 9404-9416 (2012).
28. Nielsen, C.B. et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894-1910 (2007).
29. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144 (2006).
30. Krek, A. et al. Combinatorial microRNA target predictions. *Nat Genet* **37**, 495-500 (2005).
31. Maragkakis, M. et al. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res* **39**, W145-148 (2011).
32. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**, 1278-1284 (2007).
33. Hausser, J. et al. MirZ: an integrated microRNA expression atlas and target prediction resource. *Nucleic Acids Res* **37**, W266-272 (2009).
34. Betel, D., Wilson, M., Gabow, A., Marks, D.S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**, D149-153 (2008).
35. Wang, X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* **14**, 1012-1017 (2008).
36. Hsu, S.-D. et al. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* **36**, D165-169 (2008).
37. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**, D152-157 (2011).
38. Barrett, T. & Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* **411**, 352-369 (2006).
39. Chintapalli, V.R., Wang, J. & Dow, J.A. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet* **39**, 715-720 (2007).
40. Kapushesky, M. et al. Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res* **38**, D690-698 (2010).
41. Wang, X. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res* **34**, 1646-1652 (2006).
42. Huang da, W. et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* **8**, R183 (2007).
43. Rappoport, N., Fromer, M., Schweiger, R. & Linial, M. PANDORA: analysis of protein and peptide sets through the hierarchical integration of annotations. *Nucleic Acids Res* **38**, W84-89 (2010).

44. Jensen, L.J. et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-416 (2009).
45. Haw, R., Hermjakob, H., D'Eustachio, P. & Stein, L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. *Proteomics* **11**, 3598-3613 (2011).

Yitzhak Friedman is a Ph.D. student in the Biological Chemistry department at The Hebrew University of Jerusalem (HUJI). Y.F. research focuses on the regulatory aspects of microRNAs in complex cellular systems. His main research is on the stress response of pancreatic β-cells with a focus of ER stress and apoptosis.

Michal Linial is a Professor of Bioinformatics, Molecular and Cellular Biology at the Hebrew University of Jerusalem. She is the director of the Israel Institute for Advanced Studies. She received her PhD (1986) and was trained at Stanford in molecular Neuroscience. She serves as a vice-president of the International Society for Computational Biology (ISCB). Her research interests cover experimental and computational technologies, mass spectrometry, host-pathogen, protein family evolution, microRNAs and the cell strategies for coping with stress conditions.