

SOFTWARE

Open Access



miREM: an expectation-maximization approach for prioritizing miRNAs associated with gene-set

Luqman Hakim Abdul Hadi^{1†}, Quy Xiao Xuan Lin^{1†}, Tri Tran Minh^{1†}, Marie Loh², Hong Kiat Ng¹, Agus Salim³, Richie Soong^{1,4*} and Touati Benoukraf^{1*} 

Abstract

Background: The knowledge of miRNAs regulating the expression of sets of mRNAs has led to novel insights into numerous and diverse cellular mechanisms. While a single miRNA may regulate many genes, one gene can be regulated by multiple miRNAs, presenting a complex relationship to model for accurate predictions.

Results: Here, we introduce miREM, a program that couples an expectation-maximization (EM) algorithm to the common approach of hypergeometric probability (HP), which improves the prediction and prioritization of miRNAs from gene-sets of interest. miREM has been made available through a web-server (<https://bioinfo-csi.nus.edu.sg/mirem2/>) that can be accessed through an intuitive graphical user interface. The program incorporates a large compendium of human/mouse miRNA-target prediction databases to enhance prediction. Users may upload their genes of interest in various formats as an input and select whether to consider non-conserved miRNAs, amongst filtering options. Results are reported in a rich graphical interface that allows users to: (i) prioritize predicted miRNAs through a scatterplot of HP *p*-values and EM scores; (ii) visualize the predicted miRNAs and corresponding genes through a heatmap; and (iii) identify and filter homologous or duplicated predictions by clustering them according to their seed sequences.

Conclusion: We tested miREM using RNAseq datasets from two single “spiked” knock-in miRNA experiments and two double knock-out miRNA experiments. miREM predicted these manipulated miRNAs as having high EM scores from the gene set signatures (i.e. top predictions for single knock-in and double knock-out miRNA experiments). Finally, we have demonstrated that miREM predictions are either similar or better than results provided by existing programs.

Keywords: Gene regulation, miRNA, Expectation-maximization

Background

microRNAs (miRNAs) are important modulators of gene expression in various biological systems, including development [1], carcinogenesis [2] and virus-host crosstalk [3]. Most mRNAs can be repressed by more than one miRNA and conversely, most miRNAs have many known and predicted mRNA targets. This direct impact of

miRNAs on a large number of mRNA species makes it a powerful biological regulator and therefore a prime candidate for studying diseases such as cancer, where miRNAs can act either as oncogenes (oncomiRs) or tumour-suppressors [4]. Gene repression by miRNAs is generally achieved by pairing between the 5' end of miRNAs, from the second to the seventh nucleotide (called the seed region), and the 3' untranslated region (UTR) of the gene target [5]. The strong effects of miRNAs on mRNAs regulation could, in theory, mean that specific miRNA-signatures are linked to certain gene expression patterns. Several databases predicting miRNA's targets based on various algorithms have been launched. Although the lack of specificity in miRNA's

*Correspondence: richiesoong@hotmail.com; benoukraf@nus.edu.sg

[†]Luqman Hakim Abdul Hadi, Quy Xiao Xuan Lin and Tri Tran Minh contributed equally to this work.

¹Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Dr, 117599 Singapore, Singapore

⁴Department of Pathology, National University of Singapore, Singapore, Singapore

Full list of author information is available at the end of the article



target predictions is a known fact [6], by taking advantage of these existing miRNA-target databases, numerous programs succeeded in highlighting the most relevant miRNAs affecting a gene expression pattern [7–10]. Indeed, since a miRNA can target several mRNAs, it is possible to compute the statistical significance of the overrepresentation of miRNA's targets within a gene-set. Most of the aforementioned programs use a standard hypergeometric probability (HP) to predict miRNAs involved in a biological process based on a compendium of miRNA-target databases. This strategy is in line with other tools used to compute the significance of functional annotations from gene-lists. However, when applied to searching miRNA-signatures from gene-lists, there is a considerable overlapping of gene targets between miRNAs leading to a number of significant signatures, including numerous potential false positive predictions [6, 11]. The source of false positives may result from miRNAs sharing a similar seed region. Consequently, it is important that programs and algorithms used are able to highlight the true positives. In contrast to current methods based on HP only, we introduce a novel strategy in complement to HP, which (i) 'weigh-down' the contribution from overlapping target genes when calculating the significance of each miRNA-signature using an expectation-maximization (EM) algorithm, a general probabilistic framework that can be used for this purpose [12]; and (ii) cluster all predicted miRNAs according to their seed region sequences for identifying "synonymous" predictions. To increase the specificity of our prediction, we also build a large compendium of miRNA's target predictions based on the most used databases. To our knowledge, the application of EM-algorithm as a probability measurement for significance in functional annotation tools has yet to be explored.

Implementation

miREM workflow

The miREM workflow is composed of the following five steps (Fig. 1):

- 1 miREM takes a list of differentially expressed genes (DEG) derived from either RNA-seq or the latest microarray platforms as input. Since the input DEG list is obtained from whole transcriptome level, miREM is inapplicable to the results derived from targeted sequencing technologies or microarray platforms targeting a specific gene panel. After receiving a gene list, miREM associates each transcript to its targeted miRNA(s) using the selected prediction databases, thus to create a list of potentially repressive miRNAs.
- 2 The hypergeometric p -value (and corrected p -value according to Benjamini-Hochberg) is determined for each unique miRNA so as to identify its enrichment significance.
- 3 If only one miRNA is found to have a significant p -value, the program stops and identifies only this significant miRNA as having an influence on the DEG.
- 4 If more than one miRNA can be identified, the program then selects miRNAs with corrected p -values below the specified threshold and subjects them to the EM-algorithm to establish the likelihood probability of each miRNA. miRNAs with the highest likelihood probabilities are most likely to have an influence on the DEG. Along with tab-delimited files, miREM results are available through a clustered heatmap of miRNA-gene interactions to ease visualization of the genes targeted by predicted miRNAs. This visualization also enables users to intuitively infer the kind of co-repression activity these predicted miRNAs play in the system.
- 5 Finally, predicted miRNAs are clustered according to their seed sequences in order to identify duplicated predictions (miRNAs sharing similar sequences).

This unique workflow, coupled with a rich combination of features (Table 1), makes miREM a better alternative when compared to existing softwares.

miREM's miRNA-target interactions database

Multiple features characterized in the miRNA-target interaction mechanism have been exploited for developing strategies to predict miRNA target genes, including (i) the seed complementarity between miRNA and mRNA strands; (ii) the free energy of the miRNA:mRNA duplex; (iii) the target site accessibility; (iv) the contribution of multiple binding sites and (v) the evolutionary conservation [13]. Many databases exploit one or several prediction strategies to provide miRNA-mRNA interaction information. In order to build a comprehensive resource to predict a miRNA-signature, we compiled those up-to-date information pooled from the latest releases of Diana [14], Miranda [15], mirDB [16], Pictar [17], PITA [18], RNA22 [19] and TargetScan [20] (Additional file 1: Table S1).

Querying miREM's database

miREM accepts Refseq, Ensembl, UCSC IDs or official gene names as input and uses Ensembl Biomart [21] to unify gene IDs. The miREM analysis interface allows interrogation of the reference databases in a flexible way, enabling users to select one or more preferred reference databases. If more than one database is selected, the user can opt to use only the miRNA-target interactions that are found to be common to all (intersections). Alternatively, users can select miRNA-target interactions that appear in

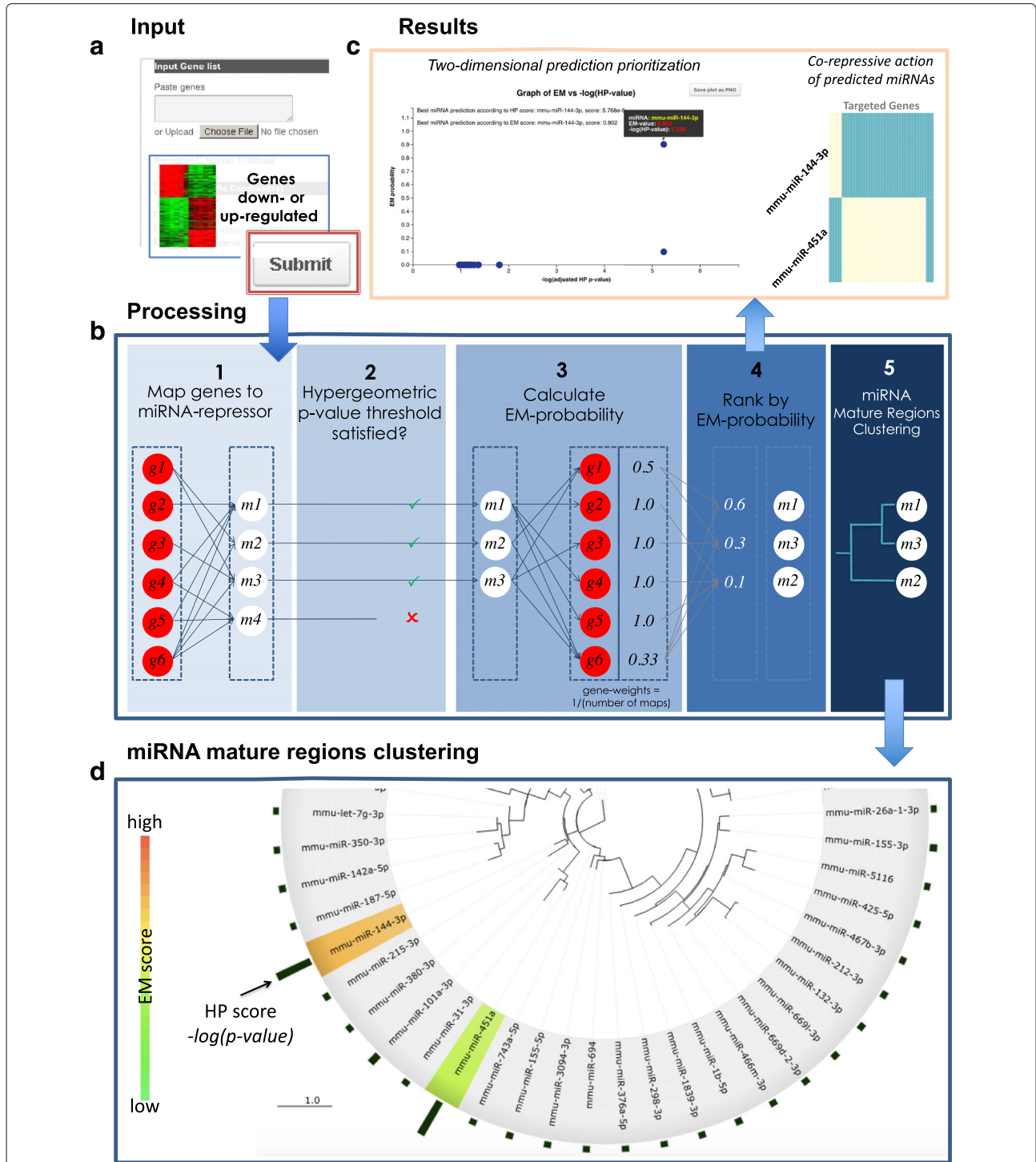


Fig. 1 miREM Workflow. **a** The gene-list and setup parameters are entered into the input page. **b** Each transcript is associated to its targeted miRNA(s) using the selected prediction databases. miRNAs are selected only if their respective HP p-values reach a pre-defined enrichment level. Then, these selected miRNAs are subjected to the EM-algorithm to establish the likelihood probability of each miRNA. miRNAs with the highest likelihood probabilities are the most likely to have an influence on the DEG. Afterwards, predicted miRNAs are clustered according to their seed region sequences in order to identify duplicated predictions. **c** Subsequently, results are displayed in a dynamic graphical interface allowing an easy data interpretation. **d** Finally, a dendrogram of miRNA seed sequences is generated to help in identifying duplicated predictions (miRNAs sharing similar sequences)

Table 1 Feature comparisons of five miRNA predicting tools

	miREM	CORNA	Geneset2miRNA	ChemiRs	Sylamer
Platform	Web-based	R package	Web-based	Web-based	Web-based standalone
Installation required	No	Yes	No	No	Optional
GUI	Yes	No	Yes	Yes	Yes
Software last update	2017	2013	2009	2015	Unspecified
Organisms supported	Human, mouse	22 species including human and mouse	6 species including human and mouse	Human	7 species including human and mouse
Reference databases	Diana, Miranda, mirDB, Pictar, PITA, RNA22, TargetScan	Miranda	mirDB, Pictar, PITA, TargetScan	Diana, Miranda, mirDB, miRWalk, RNA22, RNAhybrid, Pictar(4way), Pictar(5way), PITA, TargetScan	Unspecified
Mirbase version	21 (June 2014)	10.0 (August 2007)	11 (April 2008)	21 (June 2014)	Unspecified
Ability to specify database(s)	Yes	No	No	Yes	No
Database last updated	2017	2007	2008	2015	Unspecified
Cross-database query	Yes	No	No	Yes	No
Conserved/ nonconserved miRNA distinction	Yes	No	No	No	No
Input format	List of DEG: text file, Gene Symbol, RefSeq, Ensembl, UCSC	List of DEG: Ensembl	List of DEG: text file, Gene Symbol, RefSeq Protein ID, RefSeq Transcript ID, Ensembl, UniGene, Entrez Gene ID, UniProt/Swiss-Prot,	List of DEG: text file, Gene Symbol, Ensembl	Full gene list ranked by fold change
Output formats	Tab-delimited file Scatterplot Heatmap Phylogenetic miRNA classification	R list	CSV	CSV	jnlp
Show target gene list	Yes	No	Yes	No	No
Algorithm	Hypergeometric Expectation Maximization	Hypergeometric Fisher's exact test Chi-square	Hypergeometric	Hypergeometric	Hypergeometric
Detection of duplicated predictions	Yes	No	No	No	No

one (union of all databases) or more of the seven databases (intersections from 2 to 7 -all- reference databases). Moreover, for Targetscan and Miranda reference databases, queries can be restricted to evolutionary conserved miRNAs. Reference databases are complementary as a large portion of predicted targets are unique to each database (Additional file 2: Figure S1). The flexibility in querying reference databases provides users a leverage to adjust for prediction specificity and sensitivity.

EM algorithm formulation

Assuming we have N genes and K miRNAs, let us define an $N \times K$ matrix Z , where $z_{ik} = 1$ if gene i is repressed by miRNA k ; otherwise, $z_{ik} = 0$. We are interested in estimating the proportion of genes repressed by the k^{th} miRNA, $p_k, k = 1, \dots, K$. By observing matrix Z , in other words, we are certain which of the genes are repressed for each miRNA, then the probabilities p_k 's can be estimated by counting the number of genes repressed by the miRNA and dividing this by the total number of genes,

$$\hat{p}_k = \frac{\sum_{i=1}^N z_{ik}}{N} \tag{1}$$

However, in reality, we do not directly observe Z . Instead, through miRNA-target prediction databases, we observe matrix $Y = (y_{ik}), i = 1, \dots, N; k = 1, \dots, K$, where $y_{ik} = 1$ if the k^{th} miRNA is predicted to target the i^{th} gene, otherwise $y_{ik} = 0$. We indicate the parameter of interest θ as (p_1, \dots, p_K) . The likelihood of the complete data (Y, Z) can be written as,

$$L(\theta|Y, Z) = \prod_{i=1}^N \prod_{k=1}^K (p_k^{z_{ik}} (1 - p_k)^{1-z_{ik}})^{y_{ik}} \tag{2}$$

Note that the above likelihood function implicates a logical assumption that $P(z_{ik} = 0 | y_{ik} = 0)$, i.e., gene i cannot be targeted by miRNA k if the prediction database does not predict this.

The log-likelihood of the complete data is taken as the logarithm of the likelihood and is given by,

$$\ell(\theta|Y, Z) = \sum_{i=1}^N \sum_{k=1}^K y_{ik} [z_{ik} \log p_k + (1 - z_{ik}) \log(1 - p_k)] \tag{3}$$

To estimate θ , we will use EM algorithm that consists of two iterative steps: an E-step where we evaluate the expected value of complete data log-likelihood, $Q(\theta|Y, Z) = E[\ell(\theta|Y, Z)]$; and an M-step where we update the parameter estimates using a current estimate $Q(\theta|Y, Z)$. The algorithm starts from an initial estimate $\theta^{(0)} = (p_k^{(0)}, k = 1, \dots, K)$, where $p_k^{(0)} = \frac{1}{K}$. In iteration m , we update $\theta^{(m)}$ in two steps:

- E-step: We update the current estimate of matrix Z as,

$$\hat{z}_{ik}^{(m)} = E[z_{ik}|Y_i, \theta^{(m-1)}] = \Pr(z_{ik} = 1|Y_i, \theta^{(m-1)}) \tag{4}$$

$$= \frac{y_{ik} p_k^{(m-1)}}{\sum_{k=1}^K y_{ik} p_k^{(m-1)}}, \forall i, k \tag{5}$$

- M-step: We update the estimate of $\theta, \hat{\theta}^m$ by updating each of its component,

$$\hat{p}_k^{(m)} = \frac{\sum_{i=1}^N \hat{z}_{ik}^m}{N}, \forall k \tag{6}$$

Intuitively, in the E-step, each k^{th} miRNA with $y_{ik} = 1$ is assigned a fraction of gene i in proportion to $p_k^{(m-1)}$, and this fraction is $\hat{z}_{ik}^{(m)}$ while the M-step updates the probability of each miRNA by averaging the number of genes that are predicted to be repressed by the miRNAs. The algorithm reaches convergence in a few iterations and return the parameter estimates \hat{p}_k , which is used to calculate the EM scores.

Using HP as a filtering step for EM-algorithm inputs

A non-negligible constraint of the EM algorithm is the running-time performance. Indeed, this algorithm is based on multiple iterations with a linear complexity. Therefore, its performance decreases with increments in the input data. Hence, we attempted to limit the number of miRNAs involved in the computation to those that are more likely to be truly significant. To achieve this, we have implemented HP as an initial filtering criterion. Here, the HP is applied to each miRNA, in order to test whether the number of predicted miRNA's targets are over-represented within the gene-set. Based on the user HP p -value threshold, miREM selects only significant miRNAs and proceeds to the EM step. Where only one miRNA signature is significantly predicted by the HP, no EM is involved.

Classification of miRNA predictions

miRNAs that share similar seed regions are likely to target similar genes. As such, these miRNAs are likely to be co-predicted. In order to identify duplicated predictions, we have implemented a module to cluster miRNAs according to their levels of homology. This classification is done by multiple alignments of miRNA seed sequences using Muscle [22], followed by the generation of a dendrogram computed by PhyML [23] and displayed by jsPhyloSVG [24].

Availability and requirements

miREM webportal is freely available and can be accessed online at <https://bioinfo-csi.nus.edu.sg/mirem2/>.

Results

We have developed miREM, an HP-EM-based program designed to predict miRNA activities from a gene list. miREM's web server incorporates a large compendium of human/mouse miRNA-target prediction databases and provides rich output results facilitating prioritization and interpretation of predicted results.

To test miREM performance, we benchmarked miREM predictions against CORNA [7], GeneSet2MiRNA [8], ChemiRs [9], and Sylamer [10] results using several datasets with known miRNA activities. These are detailed in three case studies as follows:

Case study 1: knock-in miRNA experiments

We used two RNAseq expression datasets from miR-155 and miR-1 knock-in experiments in U2OS cells, respectively [25]. In these experiments, we used a gene-set of repressed genes as input (Additional file 3: Table S2) and ran miREM, CORNA, GeneSet2MiRNA and ChemiRs (Table 2 and Additional file 4: Table S3; for Sylamer, whole gene list ranked by fold change was input). miREM has predicted involving miRNAs correctly, with hsa-miR-155-5p and hsa-miR-1-3p ranked at the first and third positions respectively. Similarly, other four tools, namely CORNA, GeneSet2MiRNA,

ChemiRs and Sylamer, showed satisfactory predicting performances whereby both miRNAs involved in the experiments were accurately identified (Table 2 and Additional file 4: Table S3).

Case study 2: double knock-out miRNA experiment

In addition, we analyzed a microarray dataset derived from a miRNA double knock-out experiment [26]. Up-regulated genes from CD71⁺/Ter119⁺/FSC^{high} bone marrow cells in miR-144/451^{-/-} mice in comparison with wild-type controls (Additional file 3: Table S2) were input into miREM, CORNA and GeneSet2MiRNA (ChemiRs was excluded from this test as it provides only human miRNA-target databases; for Sylamer, the input gene list was whole gene list ranked by fold change). miREM ranked mmu-miR-144-3p and mmu-miR-451a in the first and second positions respectively. However, no prediction results were available for CORNA and GeneSet2MiRNA when *p*-value threshold was set as 0.01; even if less stringent *p*-value filtering was applied (*p*-value threshold equals to 0.05), only mmu-miR-144 was predicted for both tools (Table 2 and Additional file 4: Table S3). For Sylamer, only mmu-miR-144 was in the result (Table 2 and Additional file 4: Table S3).

Case study 3: double knock-out miRNA experiment

Finally, we compared miREM's prediction performance in a miR-181a1/181b1 double knock-out experiment [27] with other tools. In miREM, loose criterion in HP filtering step (*p*-value threshold = 0.01) resulted in a

Table 2 Performance of five miRNA prediction tools using two single miRNA knock-in and one miR-144/451 double knock-out experiments

Input data			Predictions				
Datasets	miRNA involved	DEG a	miREM (based on EM) ^b	ChemiRs ^c	GeneSet2MiRNA ^c	CORNA ^c	Sylamer ^d
Cytoplasmic RNA-seq in U2OS cells	hsa-miR-155 knock in	647	hsa-miR-155-5p (1/160)	hsa-miR-155-5p (1/118)	HSA-MIR-155 (1/23)	hsa-mir-155 (1/2)	has-miR-155
Cytoplasmic RNA-seq in U2OS cells	hsa-miR-1 knock in	743	hsa-miR-1-3p (3/9)	hsa-miR-1-3p (1/65)	HSA-MIR-1 (1/6)	hsa-mir-1 (2/4)	has-miR-1
Microarray in mice	miR-144/451 knock out	396	mmu-miR-144-3p (1/2) mmu-miR-451a (2/2)	Not Applicable ^e	mmu-mir-144 (3/3) *	mmu-mir-144 (2/4) *	mmu-miR-144

^aDEG list not applicable for Sylamer where a full gene list ranked by fold change was input

^bSettings: *p*-value threshold = 0.01, EM convergence parameter = 0.001, common mappings from 3 or more databases and nonconserved miRNAs not included

^cSettings: *p*-value threshold = 0.01 (for ChemiRs, the minimum number of databases is 5 out of 10)

^dRanking number / full prediction result number is not available in Sylamer

^eMouse databases are not provided

**p*-value threshold = 0.05 (no result with *p*-value threshold = 0.01)

Table 3 Performance of five miRNA prediction tools using a miR-181a1/b1 double knock-out experiment

Input data			Predictions				
Datasets	miRNA involved	Gene list	miREM (based on EM) *	ChemiRs	GeneSet2MiRNA **	CORNA#	Sylamer
RNA-seq in mice	miR-181a/b knock out	243	mmu-miR-181a-5p (rank 4 out of 4)	Not applicable	MMU-MIR-181B (rank 1 out of 9)	No result	No result
			mmu-miR-181b-5p (rank 1 out of 4)				

* Settings: p -value threshold = 0.0001, EM convergence parameter = 0.001, common mappings from 4 or more databases and non-conserved miRNAs not included

** p -value threshold = 0.01

p -value threshold = 0.05

long list of miRNA candidates. Hence, a more stringent screening setting was applied in HP step (p -value threshold = 0.0001). Again, miREM is the only tool which is able to predict both involving miRNAs, with mmu-miR-181b-5p and mmu-miR-181a-5p ranked in first and fourth positions out of total four predictions respectively. Nonetheless, only mmu-miR-181b was predicted by GeneSet2MiRNA while no results were obtained for CORNA and Sylamer (Table 3 and Additional file 4: Table S3).

Impact of miRNA databases vs algorithms

Integration of EM algorithm with HP test contributes to miREM's better prediction performance. In order to test the algorithm implemented in miREM while ruling out database bias, we compared the performances of miREM and CORNA (the rest of the test programs were web-server based and could not be modified) on miR-144/451 double knock-out assay using same miRNA-mRNA interaction database (miRanda mouse conserved miRNA database August 2010 release). miREM ranked miR-144 and miR-451 as first and second predictions respectively in its result, whereas 17 miRNAs were predicted by CORNA, among which miR-144 was ranked first while miR-451 ranked 5th (Additional file 5: Table S4). Hence, introduction of EM into miRNA prediction allowed us to better rank candidate miRNAs. Furthermore, the prediction result by miREM was relatively robust. We tested miREM's performances using different HP p -value thresholds and EM convergence parameters given the down-regulated gene list from hsa-miR-155 knock-in experiment. hsa-miR-155-5p remained the first-ranked candidate in various prediction settings (Additional file 6: Table S5).

Conclusion

The combination of HP and EM algorithm coupled with a large miRNA-target compendium of databases makes miREM a tool of choice to predict and prioritize miRNAs from a given gene list. Programs like miREM rely on miRNA databases, which can be a source of bias, particularly for uncommon miRNAs processed by the Ago2 endonuclease [28], an alternative

mechanism independent to Dicer. Therefore, there is still room for improving target predictions by a better characterization of miRNA targets. Overall, we have demonstrated that miREM's prediction performance is either similar or better than existing programs such as CORNA, GeneSet2MiRNA, ChemiRs and Sylamer.

Finally, the versatility of the miREM web server makes it accessible to a large panel of users including non-bioinformaticians, by facilitating result exploration and interpretation through numerous representations and a dynamic graphical interface.

Additional files

Additional file 1: Table S1. Release notes of human and mouse miRNA reference databases in miREM. (XLSX 35 kb)

Additional file 2: Figure S1. Overlap of miRNA-mRNA predicted interactions across human and mouse reference databases. (PDF 833 kb)

Additional file 3: Table S2. List of genes differentially expressed used in the case studies. (XLSX 75 kb)

Additional file 4: Table S3. Full predicted results for four tested experiments. (XLSX 514 kb)

Additional file 5: Table S4. Predicted results of miREM and CORNA with same database. (XLSX 47 kb)

Additional file 6: Table S5. Predicted results of miREM using different HP p -values and EM parameters. (XLSX 115 kb)

Abbreviations

DEG: Differentially expressed genes; EM: Expectation-maximization; HP: Hypergeometric probability; miRNA: Micro RNA; mRNA: Messenger RNA

Acknowledgements

The authors thank Sarawut Wongphayak and Von Bing Yap for their assistance and in initial testing, as well as Celestina Chin Ai Qi and Xue Yan Yam for their proofreading.

Funding

This research is supported by the RNA Biology Center at the Cancer Science Institute of Singapore, NUS, as part of funding under the Singapore Ministry of Education's Tier 3 grants, grant number MOE2014-T3-1-006 and by the Biomedical Research Council (BMRC) 01/9/21/19/618.

Availability of data and materials

Test datasets and miREM webportal are available online at <https://bioinfo-csi.nus.edu.sg/mirem2/>.

Authors' contributions

TB, LHLA and RS designed the software. TB, LHLA, TMT and QXXL coded the program. TB, LHLA, ML, AS and QXXL performed the statistical test. TB, TMT, QXXL and HKN did benchmarking. TB, LHLA, QXXL and RS wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Dr, 117599 Singapore, Singapore. ²Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore, Singapore. ³Department of Mathematics and Statistics, School of Engineering and Mathematical Sciences, La Trobe University, Bundoora, Victoria, Australia. ⁴Department of Pathology, National University of Singapore, Singapore, Singapore.

Received: 22 October 2017 Accepted: 19 July 2018

Published online: 10 August 2018

References

- Stepicheva NA, Song JL. Function and regulation of microRNA-31 in development and disease. *Mol Reprod Dev.* 2016. <https://doi.org/10.1002/mrd.22678>.
- Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: Biomarkers, functions and therapy. *Trends Mol Med.* 2014;20(8):460–9. <https://doi.org/10.1016/j.molmed.2014.06.005>.
- Powdrill MH, Desrochers GF, Singaravelu R, Pezacki JP. The role of microRNAs in metabolic interactions between viruses and their hosts. *Curr Opin Virol.* 2016;19:71–76. <https://doi.org/10.1016/j.coviro.2016.07.005>.
- Svoronos AA, Engelman DM, Slack FJ. OncomiR or Tumor Suppressor? The Duplicity of MicroRNAs in Cancer. *Cancer Res.* 2016;76(13):3666–70. <https://doi.org/10.1158/0008-5472.CAN-16-0359>.
- Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. *Genes Dev.* 2004;18(5):504–11. <https://doi.org/10.1101/gad.1184404>.
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009;136(2):215–33. <https://doi.org/10.1016/j.cell.2009.01.002>.
- Wu X, Watson M. CORNA: Testing gene lists for regulation by microRNAs. *Bioinformatics.* 2009;25(6):832–3. <https://doi.org/10.1093/bioinformatics/btp059>.
- Antonov AV, Dietmann S, Wong P, Lutter D, Mewes HW. GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res.* 2009;37(Web Server issue):323–8. <https://doi.org/10.1093/nar/gkp313>.
- Su EC, Chen YS, Tien YC, Liu J, Ho BC, Yu SL, Singh S. ChemiRs: a web application for microRNAs and chemicals. *BMC Bioinformatics.* 2016;17:167. <https://doi.org/10.1186/s12859-016-1002-0>.
- van Dongen S, Abreu-Goodger C, Enright AJ. Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods.* 2008;5(12):1023–5. <https://doi.org/10.1038/nmeth.1267>.
- Zheng H, Fu R, Wang J-T, Liu Q, Chen H, Jiang S-W. Advances in the techniques for the prediction of microRNA targets. *Int J Mol Sci.* 2013;14(4):8179–87. <https://doi.org/10.3390/ijms14048179>.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B.* 1977;39(1):1–38.
- Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD. Bioinformatic tools for microRNA dissection. *Nucleic Acids Res.* 2016;44(1):24–44. <https://doi.org/10.1093/nar/gkv1221>.
- Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* 2013;41(Web Server issue):169–73. <https://doi.org/10.1093/nar/gkt393>.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 2008;36(Database issue):149–53. <https://doi.org/10.1093/nar/gkm995>.
- Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 2015;43(Database issue):146–52. <https://doi.org/10.1093/nar/gku1104>.
- Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. *Nat Genet.* 2005;37(5):495–500. <https://doi.org/10.1038/ng1536>.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet.* 2007;39(10):1278–84. <https://doi.org/10.1038/ng2135>.
- Loher P, Rigoutsos I. Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics (Oxford, England).* 2012;28(24):3322–3. <https://doi.org/10.1093/bioinformatics/bts615>.
- Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell.* 2007;27(1):91–105. <https://doi.org/10.1016/j.molcel.2007.06.017>.
- Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kähäri AK, Keenan S, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Overduin B, Parker A, Patricio M, Pery E, Pignatelli M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Aken BL, Birney E, Harrow J, Kinsella R, Muffato M, Ruffier M, Searle SMJ, Spudich G, Trevanion SJ, Yates A, Zerbino DR, Flicke P. Ensembl 2015. *Nucleic Acids Res.* 2015;43(Database issue):662–9. <https://doi.org/10.1093/nar/gku1010>.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
- Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 2005;33(suppl 2):557–9. <https://doi.org/10.1093/nar/gki352>.
- Smits SA, Ouverney CC. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PLoS ONE.* 2010;5(8):12267. <https://doi.org/10.1371/journal.pone.0012267>.
- Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, hao Hsu S, Ghoshal K, Villén J, Bartel DP. MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell.* 2014;56(1):104–15. <https://doi.org/10.1016/j.molcel.2014.08.028>. NIHMS150003.
- Yu D, Dos Santos CO, Zhao G, Jiang J, Amigo JD, Khandros E, Dore LC, Yao Y, D'Souza J, Zhang Z, Ghaffari S, Choi J, Friend S, Tong W, Orange JS, Paw BH, Weiss MJ. miR-451 protects against erythroid oxidant stress by repressing 14-3-3zeta. *Genes Dev.* 2010;24(15):1620–33. <https://doi.org/10.1101/gad.1942110>.
- Henao-Mejia J, Williams A, Goff L, Staron M, Licona-Limón P, Kaech S, Nakayama M, Rinn J, Flavell R. The MicroRNA miR-181 Is a Critical Cellular Metabolic Rheostat Essential for NKT Cell Ontogenesis and Lymphocyte Development and Homeostasis. *Immunity.* 2013;38(5):984–97. <https://doi.org/10.1016/j.immuni.2013.02.021>. NIHMS150003.
- Cheloufi S, Dos Santos CO, Chong MMW, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature.* 2010;465(7298):584–9. <https://doi.org/10.1038/nature09092>.