

Towards a Comprehensive Understanding of miRNA Regulome and miRNA Interaction Networks

Joseph J Nalluri^{1*}, Debmalya Barh², Vasco Azevedo³ and Preetam Ghosh¹

¹Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA

²Center for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology, Purba Medinipur, West Bengal, 721172, India

³The Department of Biology General, Universidade Federal de Minas Gerais, CP 486, eMolecular Celular, Brazil

*Corresponding author: Joseph J Nalluri, Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA, Tel: 18185713728; E-mail: nallurijj@mymail.vcu.edu

Received date: August 02, 2016; Accepted date: August 09, 2016; Published date: August 16, 2016

Copyright: © 2016 Nalluri JJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

miRNAs are vital regulators of post transcriptional gene expression. Deregulation of miRNAs lead to susceptibility towards many diseases, especially cancers. Their complicated molecular mechanisms comprising of upstream transcriptions factors, downstream targets, functional and biological processes, and disease regulations are not fully comprehended yet. Hence, understanding the miRNA regulatory network comprehensively is pivotal to modulate its functions and develop miRNA therapeutics. At present, several independent databases and tools containing specific information about parts of a miRNA regulome exist in silo which prevents a holistic understanding of miRNA's molecular mechanism. Hence, integration of all scattered datasets in a cohesive manner in order to get an overarching understanding of the contributory influence and the effluence from the machinery of a miRNA regulome is critical. In this article, we present a case-report on miRegulome, a first comprehensive integrated knowledge base of miRNA regulome and miRNA interaction network analytic tools. miRegulome integrates the essential molecular modules of a miRNA regulome into a cohesive platform. We also discuss miRNA-disease interactions from miRegulome and devise graph theoretical strategies to analyze them. We also present a next-level design for an enhanced database repository for comprehensive data analysis collating diverse datasets related to miRNA biology; and present the need and challenges for the development of novel algorithms to predict new interactions between miRNAs, genes, transcription factors and diseases.

Keywords: miRNA regulome; miRNA integrated database; miRNA disease network analysis

Introduction

MicroRNAs (miRNAs) are non-coding RNA molecules which are about 22 nucleotides in size. They inhibit the expression of a target mRNA molecule by binding to its 3'-UTR through complimentary-base pairing [1]. Though they primarily act as negative regulators of gene expression [2], they have also been found to act as positive regulators [3]. A maturely transcribed miRNA possesses a mechanism for feed-back and feed-forward loop regulation through which it can regulate its own transcription process or other gene expressions significantly [4]. Thus it can effectively regulate post-transcriptional gene expression by targeting certain mRNA/s by which it modulates several signaling pathways, biological processes and pathophysiological conditions. miRNA expression is inversely proportional to its target mRNA expression. A single miRNA can target 200 mRNAs [5] and subsequently may regulate various essential biological processes such as development, aging, immunity and autoimmunity. De-regulations of miRNAs have been evidenced to associate with several types of cancers, neuronal diseases and metabolic disorders. Owing to these reasons, there has been a significant interest in miRNA regulomics and miRNA therapeutics in the bio-medical research, mainly because

of its relevance in the development of diagnostic, prognostic and therapeutic strategies.

Motivation

The challenge in deciphering and understanding the regulatory function of a miRNA is twofold. Firstly, there are several contributing elements towards a regulation of a miRNA, whose impacts are not entirely known or exhaustively comprehended yet. Most of the contributing elements have an indirect way of regulating a miRNA which makes the understanding of the regulatory network, a complex scenario. This is described in the Overview section of Figure 1. The regulatory network of a miRNA can be categorized essentially into the components-upstream regulators, downstream targets, post and post-effect modules, as show in Figure 1A miRNA expression is directly/indirectly influenced by several environmental factors, xenobiotics, chemicals and drugs.

These factors along with transcription factors (TFs) regulate the expression of a miRNA which consequently regulates the expression of its target mRNA. The downstream module contains target genes which regulate many pathways and biological functions. De-regulated pathways and biological functions cause pathophysiological disorders and diseases.

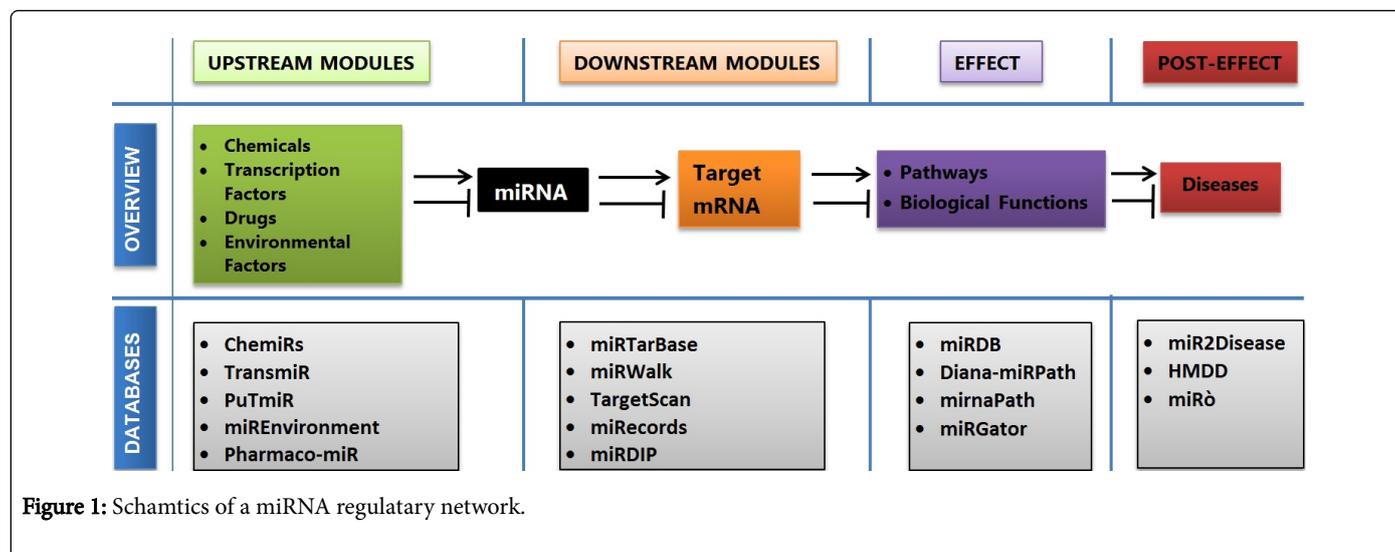


Figure 1: Schematic of a miRNA regulatory network.

Secondly, all the datasets and information pertaining to these modules are scattered across various independent studies and databases/tools which are in silo and hence disconnected with each other. Tying all this information in a cohesive manner such that, a user gets an overall picture of the contributory influence and the effluence from the machinery of a miRNA regulome is critical for understanding the regulatory function of a miRNA regulome network.

The challenge herein lies in the fact that the data in several databases and tools containing information on certain specific aspects of a miRNA need to be integrated in an intelligent manner. Figure 1 shows the isolated databases/tools which contain the specific datasets of miRNA information in the Databases section.

Upstream modules

1. Chemicals: ChemiRs [6] contains interactions for chemical, miRNA and pathways.

2. Transcription factors: TransmiR[7] and PuTmiR[8] databases contain interactions of upstream TFs and their regulated miRNAs.

3. Environmental factors: miREnvironment [9] contains the information about the environmental factors regulating miRNAs.

4. Drugs: Pharmaco-miR[10] contains interactions between drugs, miRNA and genes and links miRNAs and drug effects.

miRBase[11] contains data on sequence and annotation repositories of miRNAs.

Downstream Modules

Target mRNA/genes: miRTarBase [12], TargetScan [13], miRecords [14], miRWalk [15] and miRDIP [16] contain documented and predicted validated targets of a miRNA molecule.

Post and post-effect modules

Pathways: Diana-miRPath [17], mirnaPath [18] and miRGator [19] provide data about miRNA-target mRNA-pathway interactions.

Biological functions: miRDB [20] is a database for miRNA target prediction and functional annotation.

Diseases: miR2Disease [21], miRo [22] and Human-miRNA Disease Database (HMDD) [23] provide information of diseases in which the miRNAs are shown to regulate.

As observed, many databases and tools (in addition to the aforementioned) are dedicated to collect and study the information pertaining to specific aspects of a miRNA regulome and hence there is no cross-talk between these individual components, although they are biologically interconnected. These tools do not integrate these various pieces of information about miRNA into an all-encompassing information and analysis repository. An overall and holistic understanding would be achieved only if all the diverse sets of data were collated together and assembled into a single database and analysis repository, in a coherent manner. This would not only help the user to mine the multiple sets of information about a miRNA's regulatory function (by leveraging the holistic approach), but would also let the user analyze the results with a measure of probability and hence some certainty towards their research findings. There have not been many tools in the past which have tried to achieve this goal, for e.g., miRWalk and miRGator provide miRNA-target-pathways information but do not contain any upstream modules.

However, miRegulome [24] sets itself as one of the first comprehensive knowledge bases of miRNA regulome which integrates these diverse curated literature/data about miRNA regulome and its regulatory network, encompassing miRNAs, TFs, target genes, biological pathways, functions, diseases and chemicals, along with experimental and predicted sets of data and built-in analysis tools with statistical metrics for assessing interactions among them. miRegulome is truly unique in this aspect.

This article is a case-report detailing the novelty, various features, utilities, case-studies and analytic tools of miRegulome in Section 2. We also discuss an exclusive miRNA-disease network analysis method based on miRegulome in Section 3. In Section 4, we address the need for a more sophisticated and larger miRNA analytics repository and propose a next-level design of a more comprehensive miRNA data analytics framework and along with it, present a set of novel algorithms.

miRegulome-a Knowledge-base of miRNA Regulomics and Analysis

miRegulome is an online miRNA data and analysis repository available on <http://bnet.egr.vcu.edu/miRegulome>, that incorporates essential miRNA regulome modules and its dynamics. It is free for academic research. The current version of miRegulome (v1.0) incorporates all the downstream target genes, upstream TFs, the diverse group of chemicals as upstream regulatory modules, the signaling pathways, biological processes, and the associated diseases. miRegulome also contains four analysis tools which provide ranked list of associations with Z-score statistical assessment for associated functions, biological pathways and diseases related to the input set of queries miRNAs, genes or diseases. miRegulome contains data pertaining to the aforementioned modules and additional datasets integrated into a single analytics platform.

Database construction and contents

miRegulome has extensively collated experimentally verified data for all the upstream modules of a miRNA regulome i.e., chemicals and TFs, and downstream modules i.e., validated targets, modulated pathways, regulated BPs and associated diseases, through manual curation of published literature indexed by 3417 PubMed articles. It contains the modules of miRNA for human, mouse, rat and other species. The details of the contents of miRegulome are described in Figure 2.

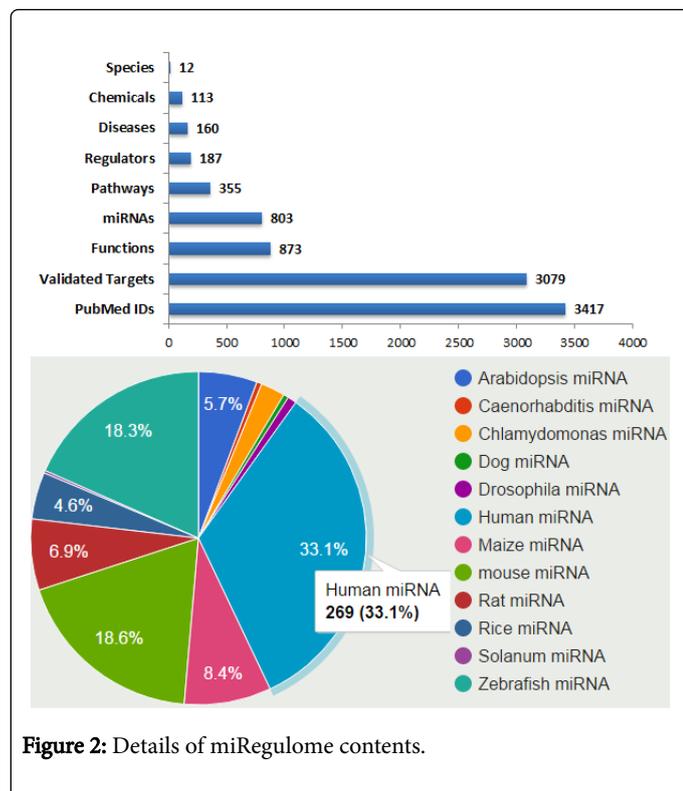


Figure 2: Details of miRegulome contents.

All the modules mentioned above are constituted in the following manner:

miRNAs and upstream chemical regulators: This module contains the information about the miRNAs which are up/down regulated in response to a chemical, drug, carcinogens, organic and inorganic

compounds, metals and other environmental factors. It also contains the species of miRNA, its expression, the experimental conditions, techniques used for detection and the PubMed ID of the chemical miRNA relationship.

Upstream TF regulators and downstream targets: Upstream TFs regulate the transcription of a miRNA, which consequently targets specific mRNA genes. This machinery is the most vital component of a miRNA regulome. Experimentally validated upstream TFs and downstream target genes/mRNAs of each miRNA that have upstream chemical regulator are manually curated and collated from the PubMed literature.

Prioritized targets and miRNA functions: The inclusion of prioritized targets and target based top miRNA functions are unique to miRegulome. Based on the number of interactions of a target in a protein-protein interaction network, a target prioritization was performed using ToppNet [25] algorithm. 11 house-keeping genes, prescribed by Eisenberg and Levanon [26] as the training set and all experimentally validated targets of each miRNA as the test set were used in the ToppNet analysis. All the targets of each miRNA were subjected to the ToppFun [25] analysis to derive their top functionalities. Also, these targets were analyzed using 'Functional Annotation' module of DAVID [27] (with default p-values cut off at 0.1) using which, the top 25 predicted functions and BPs are listed.

miRNA involved pathways: miRNAs regulate pathways significantly. This information is captured by subjecting all the validated targets of each miRNA to DAVID [27] for enrichment into Kyoto Encyclopedia of Genes and Genomes (KEGG) [28] pathways. The top ten enriched pathways which are hyperlinked to their corresponding pathways in miRNAPath [18] database for further details.

Disease module: The miRNA disease module is a vital feature of miRegulome because it essentially allows the user to explore the ways in which a miRNA is affecting a pathophysiological process. miRNA disease associations were curated along with up/down regulation of the miRNA in the disease conditions from the PubMed published literature, for the miRNAs that respond to chemical stimulus. This association is further hyperlinked to the miR2Disease [21] database for more details.

Visualization: An intuitive schematic visualization interface is developed to capture the incorporation of the above modules to display the cohesiveness of the miRNA data. Upon selection of a certain miRNA, its entire regulome is visualized with chemicals, upstream activators and repressors, validated targets, enriched top targets, pathways, function and dis-eases along with their corresponding relationships with the miRNA. The visualization also depicts the type of relationship i.e. activation, inhibition of an association. This complex interaction map is intuitive and easily interpretable as shown in Figure 3.

Utility: miRegulome's extensive data repository and advanced analytic tools can be used to test and study various biological hypothesis and pathophysiological conditions.

Using miRegulome, it can be found that hsa-mir-27b is down-regulated and hsa-mir-143 is up-regulated in obesity (Figure 4). Further-more, using pathway analysis, it can be established that hsa-mir-27b is involved in adipocytokine, insulin, and type-2 diabetes pathways and hsa-mir-143 plays an active role in lipid metabolism pathway. These pathways are significant events in obesity and

therefore, deregulation of hsa-mir-27b and hsa-mir-143 may affect these pathways and may eventually lead to obesity and diabetes. In addition, the database also provides correlation of 'obesity-mir-27b-Ribavirin' and 'obesity-mir-143-Benzo[a]pyrene'. As per miRegulome, Benzo(a)pyrene (BaP) and Ribavirin up-regulates mmu-mir-143 and has-mir-27b, respectively. It has also been known that higher Body Mass Index (BMI) lowers bio-availability of Ribavirin and causes treatment failure in obese HCV patients [29] and at the same time Benzo[a]pyrene can induce obesity [30]. In summary, it can therefore be implicated that, (a) Benzo[a]pyrene up regulates mir-143 and affects

lipid metabolism to induce obesity and (b) An aberrant expression of mir-27b may play a role in obesity-associated insulin resistance by modulating adipocytokines and Ribavirin resistance in obese patients. Similarly, it can also be assumed that these two miRNAs interlink obesity with diabetes at a new and deeper molecular level, justifying deeper investigation. Therefore, miRegulome may play an important role in exploring novel molecular mechanism behind a disease. All the results and analysis of miRegulome are supported by PubMed literature corroboration.

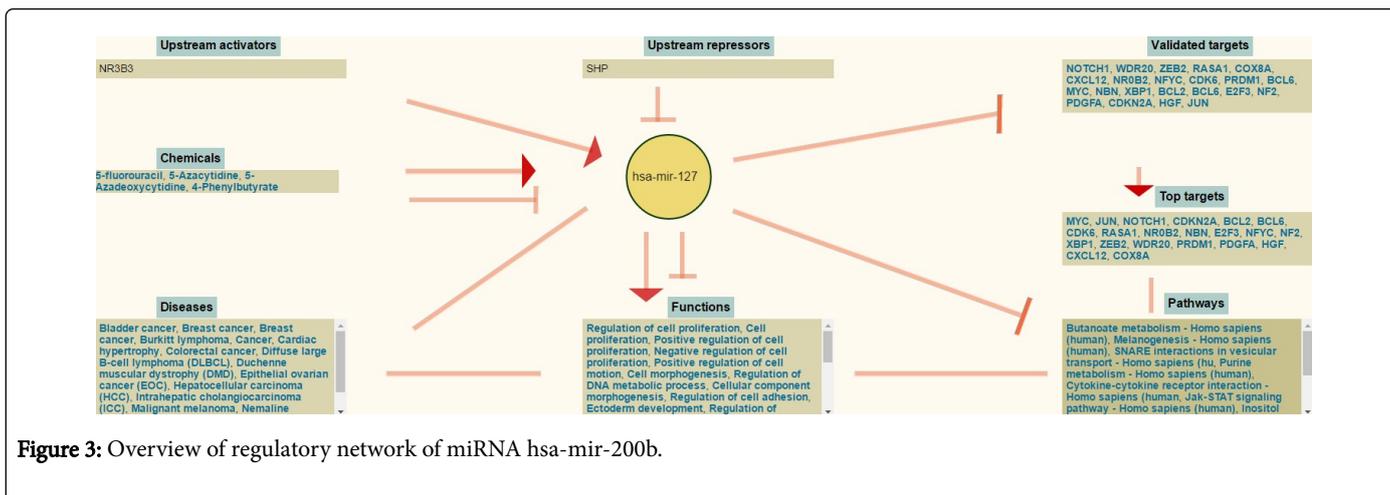


Figure 3: Overview of regulatory network of miRNA hsa-mir-200b.

miRNA interaction analysis tools

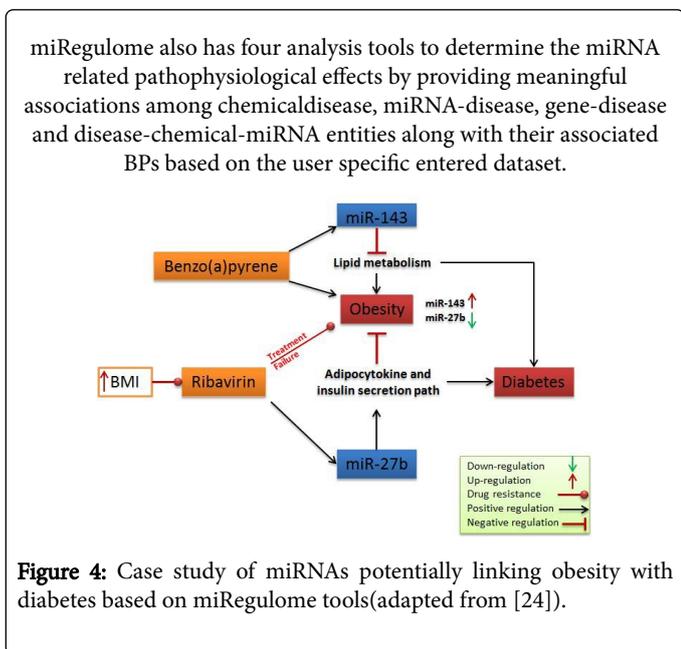


Figure 4: Case study of miRNAs potentially linking obesity with diabetes based on miRegulome tools(adapted from [24]).

These tools do not assert a direct relationship between the entities but highlight the top results by which the user can explore and test their hypothesis for indirect/direct associations between them. The tools are:

Chemical-disease analysis: Upon selecting a certain chemical, the tool queries all the miRNAs regulated with the chemical in the database, after which all the diseases in which these observed miRNAs are regulated are retrieved; thereby depicting an indirect association of

chemical onto diseases. The tool displays their disease names, their count of associations (number of PubMed IDs citing it) as recorded in the database and their respective Z-scores, giving a statistical significance of the obtained results. The tool also displays the BPs associated with the miRNAs with their count of associations, thereby giving a larger context of the chemical-miRNA relationship effecting biological processes.

miRNA-disease analysis: Upon entering a set of miRNAs, the tool provides three sets of data for the user to get a comprehensive understanding of the results. All the diseases related to the input miRNAs are retrieved and the diseases are ranked as per the maximum number of recorded PubMed IDs citing the miRNA-disease association and displays them. The top diseases are listed with their corresponding PubMeds and the respective individual up/down regulations between the input miRNAs and diseases. The tools also display Z-scores for all diseases presenting its statistical significance in the available miRegulome repository. Using this data, a user can not only observe the cumulative effect of input miRNAs on the diseases but also the impact of each one in the disease. Furthermore, the tool displays the top BPs associated with the input miRNAs.

Gene-disease analysis: Upon entering a list of input genes, the tool searches for all the miRNAs associated with the set of genes and counts the number of gene-miRNA PubMed indexed association counts. Thereafter, the tool searches and counts the association counts (PubMed IDs) between the observed miRNAs and diseases. These diseases are ranked as per the maximum number of relationships i.e., PubMed entries found in the database. Similarly, the tool also displays the top BPs which are associated with the observed miRNAs and the set of input genes.

Disease: miRNA/Chemical analysis: This tool works conversely to the Chemical-miRNA analysis tool. After a set of diseases are entered,

the tool retrieves all regulated miRNAs pertaining to the diseases and subsequently retrieves all the chemicals which regulate the retrieved miRNAs. This provides an insight into the possible role of chemicals in miRNA regulation which are deregulated in the set of diseases entered.

Salient features

End-to-end modular understanding of miRNA regulatory network: Considering Figure 3 and related data from miRegulome, it can be construed in Figure 5 that hsa-miR-200b:

- Forms a feed-back loop with TGFβ1,
- P53 activates and TGFβ1 inhibits its expression,
- Inhibits MAPK, WNT, and AKT signalling pathways,
- Inhibits cell cycle and cell proliferation by targeting cell cycle regulators and oncogenes like CCND1 VEGFA,MYC, NOTCH1, MET, EGFR etc.
- Is involved in cancer associated pathways,
- Is down-regulated in response to arsenic carcinogen,
- Is up-regulated by chemotherapy drug Gemcitabine and down-regulated in Docetaxel resistant cancers, and
- Is down-regulated in several cancers.

Therefore, it may be implicated that hsa-miR-200b could be a tumor suppressor miRNA and may be a potential therapeutic for a wide range of cancers.

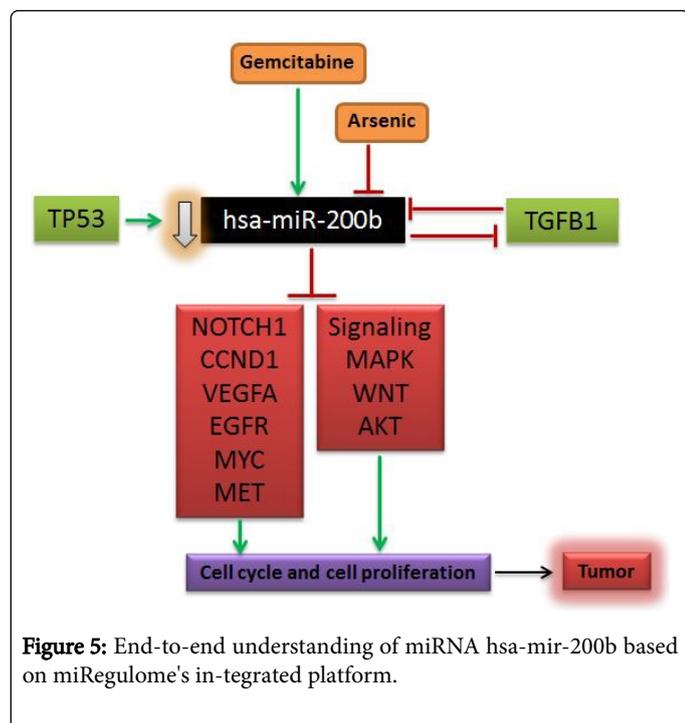


Figure 5: End-to-end understanding of miRNA hsa-mir-200b based on miRegulome's in-tegrated platform.

The above understanding regarding hsa-mir-200b is only possible because of the coherent integration of the several data-sets related to hsa-mir-200b.

miRNA interaction networks: miRegulome's unique integrated platform provides users with extensive miRNA regulatory networks, namely:

1. TF-miRNA- enriched top target network
2. Chemical/Drug-miRNA- disease network

3. TF-miRNA-enriched targets- pathways network

Based on the miRegulome's extensive knowledge base, it can also be rich data bank for the development of further novel tools, and algorithms to study specific aspects of a miRNA regulome, in-depth. The next section details a graph theoretical approach and analyses.

miRNA-Diseases Interaction Networks from miRegulome

Multi-level interactions of miRNA and diseases are a complex web of interactions, considering the fact that a miRNA regulates upto 50 diseases and targets upto 200 mRNA molecules. There have been several studies identifying and predicting miRNA-diseases associations [31,32].

miRegulome has an extensive collection of miRNA-disease interactions curated from literature, which can be studied via the application of net-work science. One such approach is maximum weighted matching model, a graph theoretical algorithm which provides the result by solving an optimization equation of determining the most prominent set of diseases. This algorithm determines and prioritizes the set of diseases which are most certainly impacted upon the activation of a group of queried miRNAs, in a miRNA disease network. This approach is implemented in a spin-off tool of miRegulome, titled DISMIRA which presents an interactive visualization feature and helps the user in exploring the networking dynamics of miRNAs and diseases. The tool also allows the users to study the miRNA disease networks of interest, by analyzing their neighbours, paths and topological features. DIS-MIRA can be accessed online for free at <http://dismira.egr.vcu.edu>.

Maximum weighted matching based analysis

As mentioned earlier, single or multiple miRNA is/are up- or down-regulated in one or a set of disease. The instances of up and down-regulations between a miRNA and disease, denote the strength of association between the pair. A bipartite graph [33] is used to map the interactions of miRNAs and diseases. A bipartite graph is a graph $G(V; E)$ in which the set of vertices V can be partitioned into two disjoint sets V_1 and V_2 such that every edge connects a vertex in V_1 to the one in V_2 [33]. In this model, miRNAs and diseases have been categorized as two disjoint sets and an edge denotes an association between them. Herein, the edges are weighted i.e., the number of publications citing up/down regulations between a miRNA-disease pair. Based on this, a weighted network consisting of miRNA-disease interactions is derived. In the graph $G(V; E)$, if there is a set of edges such that no two edges share a common end vertex, it is known as a matching. Maximum matching is a matching which has the largest possible set of edges. A maximum weighted matching (MWM) is a maximum matching in which the sum of the weights of the edges is maximum. The application of MWM on miRNA disease network provides us the strongest miRNA-disease pair combinations given a set of active miRNAs. The results give the cumulative impact of a set of activated miRNAs on the set of associated diseases, which are most certainly impacted. The goal is primarily to present a concise list of diseases with highest confidence of being influenced and to present an association between a set of miRNAs onto a set of diseases. This is vital to bear because miRNAs and diseases tend to interact closely in sets and groups and hence a tool in prioritizing disease candidates is helpful in presenting a comprehensive and yet concise list, displaying the cumulative impact of specified miRNAs.

This section detailed an example of a strategy and algorithm that is developed to study miRNA disease interaction networks based on miRegulome. Similarly, miRegulome can also be used to develop tools to study miRNA-gene TF networks, miRNA drug and miRNA gene networks.

Novel Database Frameworks and Algorithms

Due to the increasing amount of information related to miRNA and its associations and functions being amassed, it is imperative for any miRNA data repository to evolve in capacity and functionality. With the rapidly increasing availability of data via high-throughput and next-generation sequencing technologies, novel tools and integrated platforms have to be designed and developed. The development of algorithms and techniques for reconstructing miRNA interaction networks (popularly also known as network inference) from large-scale experimental data sets has been the current focus of the research community. Apart from network inference models, there is also a need for data driven deep curation approach. A deep curation approach consists of creating a molecular-level interaction map via large-scale integration of information, such as literature, other databases and high-throughput data [34]. Upon this molecular map, users can apply their own hypothesis and evaluate the findings. Such strategies can be implemented on a comprehensive integrated miRNA analytics platform and hence, there is a dire need for such a collated database frameworks and models.

Collated database

The challenge of a collated database is not only to integrate common sets of information into a single repository but also to bind them with relevance so as to provide a comprehensive and intricate working model of miRNA related biology. Here are few examples of the types of diverse sets of information about miRNAs currently available:

- miRNA target predictions
- miRNA disease predictions
- miRNA disease with expression scores
- Predicted disease-specific miRNA-miRNA interaction networks
- miRNA caused DNA methylation
- miRNA from different species such as arabidopsis, caenorhabditis, chlamydomonas, dog, drosophila, maize, rice, solanum and zebrafish
- miRNA and epigenetic associations
- New miRNA and drug interactions and results

Bearing this in mind, a framework of a collated database has to be developed combining extensive miRNA regulomics data both experimental and predicted. This database ought to incorporate diverse sets of data which are not only substantial in detail but also in variety of sources, modules and functionalities. A palpable example of such a collated database described above can be envisioned in the following example—a database that will comprise of the following modules.

- **miRegulome:** constituting of the essential modules of miRNA regulome i.e. upstream regulators, downstream targets, validated targets, affected functional and biological processes, and disease regulations pertaining to various species.
- **PhenomiR[35]:** constituting miRNA expression data sets with regulation with diseases. This database contains differentially

regulated miRNA expression data in diseases and biological processes.

- **miREnvironment:** gives the information about the phenotypes being affected when environmental factors affects the miRNAs.
- **Pharmaco-miR:** captures the interactions between miRNAs, genes, and drugs. This information adeptly compliments the miRNA-target gene interactions recorded in the miRegulome database.
- **EpimiR[36]:** This data source contains the interactions and information between the epigenetic modification and miRNAs in the context of several diseases. It also provides information about the predicted transcription start sites which will help in providing more details in miRNA guided post-transcriptional gene regulation.
- **miRsig [37]:** This database contains predicted networks of disease-specific miRNA-miRNA interactions based on network inference strategies. This tool uses the miRNA-disease interactions from PhenomiR along with their expression scores.

A schematic overview of this collated database is represented in Figure 6. It can be observed that miRNA related data repository is not only huge but also complex. There are many direct and indirect associations between the datasets. This model presents a merger of six major databases. The data consisting in them are diverse, overlapping and in some cases, complementing each other.

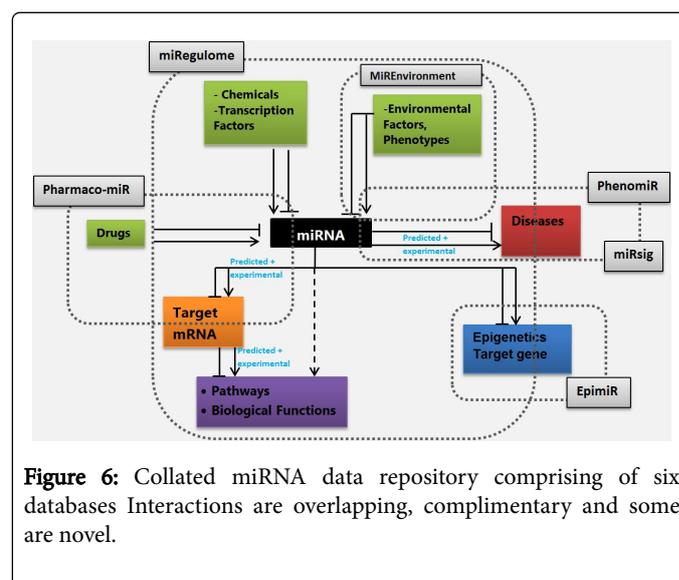


Figure 6: Collated miRNA data repository comprising of six databases Interactions are overlapping, complimentary and some are novel.

Algorithms

The challenges with a data repository of this nature is not only data handling, integrating and updating the data and making the database scalable but also to be able to devise novel analytic tools that answer important biological questions. Some of the data sources contain experimentally validated information about miRNAs, diseases, genes and TFs, while other data sources contain predicted associations between these entities. The ultimate aim is to derive and predict associations between these entities with a significant certainty. Thus, algorithms have to be conceived bearing on the existing experimentally validated associations available and be able to predict new undiscovered associations between the entities. In the approach mentioned in Section 3.1, the miRNA disease network derived from the miRegulome was modelled as a weighted graph. Herein, the weight

of the edges/associations was the count of PubMed IDs citing the association. On the contrary, in the weighted miRNA-disease network model derived from PhenomiR, the edge weights would be experimentally validated fold change expression scores of regulation between miRNA and diseases. The nature of edge weights is different in both the scenarios for the same entities. The former approach records and capitalizes on the available literature while the latter uses the fold change expression scores from experiments. To tie these multiple sources of information into a holistic model requires defining new metrics for edge scores. A miRNA disease edge can possess different edge scores based on either literature count or experimental or predicted information. Such collation of diverse sets of information between the same entities requires conception of novel algorithms and approaches in deciphering the patterns. Also, there are several network inference algorithms which have been extensively deployed to predict novel associations between network models of biological entities. The usage of network inference algorithms in the DREAM challenge, to reconstruct the gene-TF regulatory network [37] is a prime example. The prediction of disease specific miRNA-miRNA interaction network via a consensus-based network inference approach [36] is a recent example. Traditionally, these network inference algorithms have used experimentally available expression datasets as input to predict new associations based on the patterns of co-expression observed in the experiments. However, to use network inference algorithms on networks which are not only constructed based on expression datasets but also from curated literature as in Section 3.1, new inference methodologies need to be conceived. Hence, collation of diverse sets of data and multiple data definitions for the same entities require further in-depth investigation for the development of novel algorithms.

Conclusion

We present a case report towards a comprehensive understanding of miRNA regulatory network using, miRegulome and its features. miRegulome is the first-of-its-kind, comprehensive miRNA knowledge-base. In this report, we detail the novelty, tools, analytics and utilities of miRegulome. We also present network based inference strategies built on miRegulome database. We also present the need to develop a novel framework of a collated database capturing diverse interactions and associations. Novel network models need to be developed between these entities and specific algorithms will have to be conceived to answer important biological questions.

References

1. Chen X, Liu MX, Yan GY (2012) RWRMDA: predicting novel human microRNA-disease associations. *Mol BioSyst* 8: 2792-2798.
2. Bartel DP (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* 116: 281-297.
3. Vasudevan S, Tong Y, Steitz JA (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318: 1931-1934.
4. Gurtan M, Sharp PA (2013) The role of miRNAs in regulating gene expression networks. *J Mol Biol* 425: 3582-3600.
5. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. *Nature Genetics* 37: 495-500.
6. Wang J, Lu M, Qiu C, Cui Q (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res* 38: D119-D122.
7. Bandyopadhyay S, Bhattacharyya M (2010) PuTmir: a database for extracting neighboring transcription factors of human microRNAs. *BMC Bioinformatics* 11: 190.
8. Su ECY, Chen YS, Tien YC, Liu J, Ho BC (2016) Chemirs: a web application for microRNAs and chemicals. *BMC Bioinformatics* 17: 1.
9. Yang Q, Qiu C, Yang J, Wu Q, Cui Q (2011) miRenvironment database: providing a bridge for microRNAs, environmental factors and phenotypes. *Bioinformatics* 27: 3329-3330.
10. Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N (2013) PharmacomiR: linking microRNAs and drug effects. *Briefings in Bioinformatics* 15: 648-659.
11. Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ (2006) mirbase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140-D144.
12. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, et al. (2010) mirtarbase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* pp: 1-7.
13. Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4: e05005.
14. Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) Mirecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 37: D105-D110.
15. Dweep H, Sticht C, Pandey P, Gretz N (2011) mirwalk-database: prediction of possible miRNA binding sites by walking the genes of three genomes. *J Biomed Inform* 44: 839-847.
16. Shirdel EA, Xie W, Mak TW, Jurisica I (2011) Navigating the microneome using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLOS ONE*.
17. Vlachos SI, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, et al. (2015) Diana-mirpath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res*.
18. Chiromatzo AO, Oliveira TYK, Pereira G, Costa AY, Montesco CAE, et al. (2007) miRNApath: a database of miRNAs, target genes and metabolic pathways. *Genet Mol Res* 6: 859-865.
19. Nam S, Kim B, Shin S, Lee S (2008) Mirgator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res* 36: D159-D164.
20. Wang X (2008) Mirdb: a microRNA target prediction and functional annotation database with a wiki interface. *Rna* 14: 1012-1017.
21. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, et al. (2009) Mir2disease: a manually curated database for mi-crona deregulation in human disease. *Nucleic Acids Res* 37: D98-D104.
22. Lagana S, Forte A, Giudice M, Arena P, Puglisi L, (2009) Miro: a miRNA knowledge base. *Database* bap008.
23. Li Y, Qiu C, Tu J, Geng B, Yang J, et al. (2013) Hmdd v2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*.
24. Barh D, Kamapantula B, Jain N, Nalluri J, Bhattacharya A, et al. (2015) Mireg-ulome: a knowledge-base of miRNA regulomics and analysis. *Scientific Reports* 5.
25. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37: W305-W311.
26. Eisenberg E, Levanon EY (2013) Human housekeeping genes, revisited. *Trends in Genetics* 29: 569-574.
27. Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols* 4: 44-57.
28. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2011) Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* gkr988.
29. Alsio A, Rembeck K, Askarieh G, Christensen PB, Farkkila M, et al. (2012) Impact of obesity on the bioavailability of peginterferon-2a and ribavirin and treatment outcome for chronic hepatitis c genotype 2 or 3. *PLOS ONE* 7: e37521.
30. Irigaray P, Newby JA, Lacomme S, Belpomme D (2007) Overweight/obesity and cancer genesis: more than a biological link. *Biomedicine & Pharmacotherapy* 61: 665-678.
31. Nalluri J, Kamapantula BK, Barh D, Jain N, Bhattacharya A, et al. (2015) Dis-mira: Prioritization of disease candidates in miRNA-disease

-
- associations based on maximum weighted matching inference model and motif-based analysis. *BMC Genomics* 16: S12.
32. Lan W, Wang J, Li M, Liu J, Wu FX, et al. (1998) Predicting microRNA-disease associations based on improved microRNA and disease similarities. *IEEE/ACM Trans Comput Biol Bioinform*.
 33. Asratian S, Denley MJ, Haggkvist R (1988) *Bipartite graphs and their applications*, no. 131, Cambridge University Press.
 34. Kaizu K, Ghosh S, Matsuoka Y, Moriya H, Shimizu-Yoshida Y, et al. (2010) A comprehensive molecular interaction map of the budding yeast cell cycle. *Molecular Systems Biology* 6: 415.
 35. Dai E, Yu X, Zhang Y, Meng F, Wang S, et al. (2014) Epimir: a database of curated mutual regulation between miRNAs and epigenetic modifications. Database bau023.
 36. Nalluri J, Barh D, Azevedo V, Ghosh P (2016) mirsig: a consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures. *Scientific Reports* (Under Review).
 37. Marbach D, Costello JC, Kuner R, Vega NM, Prill RJ, et al. (2012) Wisdom of crowds for robust gene network inference. *Nature Methods* 9: 796-804.