2017

NETWORK ANALYTICS FOR THE MIRNA REGULOME AND MIRNA-DISEASE INTERACTIONS

Joseph Jayakar Nalluri
Virginia Commonwealth University

DISSERTATION

NETWORK ANALYTICS FOR THE MIRNA REGULOME AND MIRNA-DISEASE INTERACTIONS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

JOSEPH J. NALLURI

B.S., Computer Science - July 2004 to July 2009

M.S., Computer Science - January 2010 to May 2012

Director: Preetam Ghosh, Ph.D.,

Associate Professor, Department of Computer Science

Committee Members:

Milos Manic, Ph.D.

Thang Dinh, Ph.D.

Devanand Sarkar, MBBS, Ph.D.

Vladimir Vladimirov, M.D., Ph.D.

Virginia Commonwealth University

Richmond, Virginia

August, 2017

Acknowledgements

Every good thing given and every perfect gift is from above, coming down from God our Father ~ James 1:17 (The Bible)

I am greatly indebted to Dr. Preetam Ghosh, my adviser for mentoring and guiding me throughout the course of my doctoral studies. I have learned a great deal from him not only through our conversations but also in my study of him, his intellect and his work ethic. There have been very few people about whom I can distinctly say, that they have positively shaped my life for better and he remains one of the strongest in that category. He has always motivated me with inspiration rather than an expectation or an obligation, and for that I will always consider him as a role mentor.

I am fortunate to have a committee of esteemed researchers who made my progress their goal and invested in me and my work, generously. I am thankful to Dr. Manic for always reminding me that science is a vast field and my work is a tiny contribution in it and to never forget the bigger picture. Dr. Dinh remains an inspiration to me because of his admirable work and intelligence. I have often aspired to produce the kind of exciting work that he does. His teachings, contributions and honest career/life advices have indeed benefited me during my tenure as a doctoral student. I am thankful to Dr. Sarkar for his suggestions, guidance and thought provoking insights that he has provided. Overall, I am thankful for his goodwill towards me. Finally, I am immensely grateful to have Dr. Vladimirov on my team. He has extensively and generously provided me with his time, guidance and data sets. I thank him for being ever present to provide in-depth explanations, directions and suggestions to better my work. His contribution has undoubtedly strengthened my

dissertation. I greatly look forward to working with him in the near future.

I am thankful to my wife, Katharina for her numerous sacrifices, perpetual understanding and for always encouraging and keeping me motivated. She has been extremely gracious. My heartfelt gratitude to my family for standing beside me throughout these years and being a testament of the beautiful things in my life, in addition to research.

Lastly, I am thankful to a great group of friends. Bhanu Kamapantula, for being a motivator at work. Pratip Rana and Syed Khajamoinuddin for being ever helpful. Mack Stump, our system administrator from whom I have a learned a great deal and he continues to be an inspiration. My great friends, outside of work, who have added a dimension of friendship and richness in my life that cannot be gainsaid. The community of *Stack Overflow* who have helped me immensely all these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

DISSERTATION

NETWORK ANALYTICS FOR THE MIRNA REGULOME AND MIRNA-DISEASE INTERACTIONS

By Joseph J. Nalluri

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2017.

Director: Preetam Ghosh, Ph.D.,
Associate Professor, Department of Computer Science

miRNAs are non-coding RNAs of ~ 22 nucleotides in length that inhibit gene expression at the post-transcriptional level. By virtue of this gene regulation mechanism, miRNAs play a critical role in several biological processes and patho-physiological conditions, including cancers. miRNA behavior is a result of a multi-level complex interaction network involving miRNA-mRNA, TF-miRNA-gene, and miRNA-chemical interactions; hence the precise patterns through which a miRNA regulates a certain disease(s) are still elusive. Herein, I have developed an integrative genomics methods/pipeline to (i) build a miRNA regulomics and data analytics repository, (ii) create/model these interactions into networks and use optimization techniques, motif based analyses, network inference strategies and influence diffusion concepts to predict miRNA regulations and its role in diseases, especially related to cancers. By these methods, we are able to determine the regulatory behavior of miRNAs and po-

tential causal miRNAs in specific diseases and potential biomarkers/targets for drug and medicinal therapeutics.

CHAPTER 1

INTRODUCTION

microRNAs (miRNAs) are small non-coding RNAs that inhibit post-transcriptional gene expression by complementary base pairing at the 3-UTRs of target messenger RNAs (mRNAs)[1]. Transcription of a miRNA coding gene is under direct control of transcription factors (TFs). Expression of a miRNA is also regulated by environmental factors, xenobiotics, and drugs. These factors essentially regulate TFs and consequently regulate transcription of miRNAs[2]. A TF can positively or negatively regulate miRNA transcription. A transcribed miRNA, by virtue of its feed-back and feed-forward loop regulation mechanisms, regulates its own transcription machinery or expression of other genes and thereby regulates gene expression. A single miRNA may target nearly 200 mRNAs[3] and henceforth may regulate multiple signaling pathways and various essential biological processes (BPs) such as development, aging, immunity, and autoimmunity, etc.

Deregulations of miRNAs are well documented for their association with various patho-physiological conditions including different types of cancers, metabolic disorders, and neuronal diseases among others. Therefore, understanding miRNA regulation is of high importance in bio-medical research. Exploration of the entire regulome of a miRNA is indispensable to understand its biology and mechanisms through which it regulates gene expression in a given biological condition. Understanding of such mechanisms will help in developing diagnostic, prognostic, and therapeutic strategies. The regulome of a miRNA essentially consists of modules such as upstream regulators, downstream targets, modulated pathways, and regulated BPs. The regulome

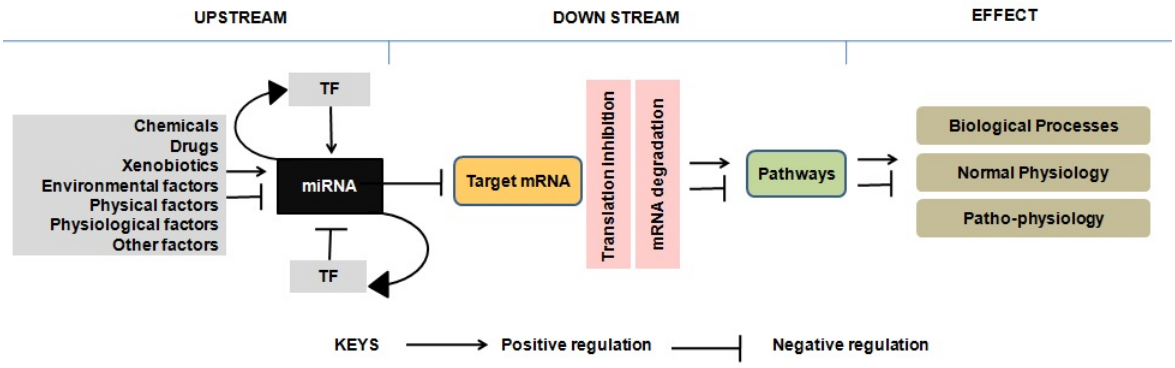


Fig. 1.: Schematic diagram of modules and their inter-relationships in a miRNA regulome

also considers associated diseases when a miRNA is deregulated. A miRNA regulome is presented in Figure 1.

Identifying and predicting miRNA and disease associations, has been extensively researched in the past few years [4–7]. However, the precise mechanisms of miRNAs regulating diseases are still unclear. Hence, gathering valuable evidence regarding identification of miRNAs influencing human diseases has become a widespread interest in arena of bio-medical research with a future towards the enhancement of human medicine. Hence, in brief, the goal of this thesis is as follows,

Goal: To determine/identify a miRNA-regulatory network using network scientific methodologies from integration of disparate data sources, especially in the context of diseases.

In this work, we try to comprehend miRNA regulatory behavior which is a result of a multi-level complex interaction network involving miRNA-mRNA, TF-miRNA-gene, miRNA-chemical and miRNA-SNP based interactions. We develop several methods and strategies to explore these aspects. In Chapter 2, we present a brief background of various studies and computational tools developed to determine these

aforementioned associations and factors.

List of contributions

1. In Chapter 3, we discuss several individual miRNA data repositories and present a necessity for an integrated platform with several analytic tools embedded into it. Thereafter, we design and create an integrated data platform for miRNA regulomics.

2. In Chapter 4, we model the miRNA-disease associations into a bipartite graph model, and determine the key set of miRNAs for a set of diseases using *Maximum Weighted Matching* optimization methodology. We also study the motif patterns observed in miRNA-disease networks and study the network topological aspects.

3. In Chapter 5, we use a consensus-based methodology to predict miRNA-miRNA interaction networks based on miRNA expression values recorded in various diseases. We also identify signature core miRNA-miRNA interaction component among multiple cancers in several categories.

4. In Chapter 6, we use *influence diffusion* theory to quantify influence diffusion in a miRNA-miRNA regulation network across several disease classes and determine critical causal miRNAs which play a causal role.

5. In Chapter 7, we determine further regulatory behavior imposed by single nucleotide polymorphism (SNP) based interactions with genes and TFs and their impact on miRNA-miRNA interaction networks of particular diseases. Herein, we predict SNP-related causal miRNAs in two diseases.

CHAPTER 2

LITERATURE REVIEW

Several existing miRNA-related databases individually provide information on specific aspects of a miRNA. For example, *miRbase*[8] maintains data on sequence repositories, *mir2Disease*[9] provides miRNA-disease relationships, *TransmiR*[10] maintains information on miRNAs and their upstream transcription factors, and *miREnvironment*[11] offers miRNA regulation in response to environmental factors. Many databases such as *miRecords*[12], *miRWalk*[13], *mirDIP*[14] , *miRTarBase*[15] etc. have been developed to enlist predicted and experimentally validated targets of miRNAs. However, none of these databases provide the entire regulome of a miRNA nor are helpful in understanding the miRNA biology or function as a stand-alone data analytics repository. Attempts have been made to understand miRNA interactome at systems level in C. elegans (TF-miRNA-TF interactions)[16] through computational simulation of miRNA regulated overall gene expression program and cross-talk between miRNA targets[17] and by constructing regulatory models of miRNA-kinase-TF, miRNA-TF, and TF-TF. Similarly, systems approach to predict TF-miRNA crosstalk in human protein interactome, demonstration of regulatory principles among miRNAs, TFs, and miRNA target genes, miRNA-mRNA and miRNA-miRNA interactions, and tissue-specific miRNA- TF regulatory networks are also have been attempted to explore the miRNA interactome. Nevertheless, these works are mostly computational predictions and do not provide the entire regulome of miRNA. Given the importance of miRNA in biomedical research, disease diagnosis, prognosis, and therapy; there has been an inflow of new miRNA related information in recent years.

Therefore, a novel data platform that provides all essential details of a miRNA regulome is necessary. Similarly, a state-of-the-art analysis platform exploring mechanisms behind various biological and patho-physiological processes which a miRNA regulates is also required. Owing to these reasons, we developed a miRNA data and analytics repository, *miRegulome*. *miRegulome* aims to address the need for such a novel database that represents the entirety of miRNA regulome. In the current version of miRegulome (v1.0) we have incorporated all the downstream modules and transcription factors (TFs) and diverse group of chemicals as the upstream regulatory modules and their correlations. *miRegulome* is explained further in-depth in Chapter 3.

For Chapter 4, the goal was to develop approaches that predict/determine associations between miRNAs and diseases. One of the preliminary works in developing miRNA-disease prediction models demonstrated that miRNAs related to same diseases tended to work together as miRNA groups[9]. This is an significant observation. It necessitates that any model of miRNA-disease association/prediction which claims to be effective considers this *homologous* nature of a miRNA. Jiang, et al., 2010 [6] uses the same approach and further derives a functional similarity between disease-related miRNAs and phenotype similarities to derive a score which evaluates the likelihood of association of a miRNA and the disease. Jiang, et al., 2010 [18] uses the disease-gene associations to develop a $Na\ddot{i}ve - Bayes$ model, which prioritizes candidate miRNAs based on their genomic distribution. This model relies heavily on the associations between gene-disease and interactions of miRNA and target. However, both these models have high false-positives and high false-negatives in their predictions [19]. This limitation was however, addressed [4], by training a support vector machine classifier based on the input set of features extracted from false-positives and false-negative predicted associations. As demonstrated by Lu, et al. 2008 [20], miRNA-set families tend to closely work towards certain diseases. Hence, implic-

itly diseases tend to affect the working of other diseases too. This has also been researched [21], where specifically prostate cancer and non-prostate cancer miRNAs are distinguished by the usage of topological features wherein, a prioritization of disease candidate was performed using a network-centric method. Apart from using disease-gene information, few models have used the assumption that miRNA loci and Online Mendelian Inheritance in Man (OMIM) disease loci may contain significant overlaps [22]. This significance score is calculated and used to identify potential associations between miRNAs and OMIM diseases. Chen, et al, 2012 [19] uses global network similarity measure as compared to local network information to implement a random walk on a functionally similar miRNA network, which prioritizes candidate miRNAs for specified diseases. Xuan, et al., 2013 [5] improvises the miRNA functionality estimated approach by appending disease phenotype similarity information and content of disease terms to the existing method. This is used to assign weight to miRNA-disease associations and a weighted k-most similar neighbor based prediction method is deployed. Global network similarity is also used in the inference methods presented [7], where apart from miRNA-similarity and phenotype-similarity inferences, a network based inference model is used. In this model [7], the miRNAs related to queried miRNA are ranked and associated with ranked disease phenotypes associated with target phenotype, thereby relying on known gene-phenotype associations. Graph theory has been extensively used to model and analyze such biological networks [20] and especially bipartite graph modeling has been used to model the miRNA-disease network [19][7][23][20]. Chen, et al., 2014 [24] has tried to overcome the limitations posed through various previous works, by developing an algorithm of Regularized Least Squares for miRNA-disease association (RLSMDA). Previous models like that of Chen, et al., 2012 [19] which although demonstrate high accuracy in prediction based on their case studies and cross-validation, cannot work in scenar-

ios where associations between the diseases and miRNAs are unknownn; and hence cannot predict novel miRNA-disease associations. Chen and Zhang [7] addressed this in their work, which could predict novel associations between diseases and miRNAs, with no prior knowledge of their association. However, its performance was inferior to that of Chen, et al. [19] based on cross-validation results [24]. The work presented by Chen, et. al [24] uses the miRNA functional similarity and disease functional similarity [25] and devises an optimization formulation to generate a continuous classification function which calculates the probablity score of each miRNA to a given disease. Using graph theory, some network inference based prediction algorithms have also been used[26]. In this case, three networks: environmental factors (EF)-miRNA, EF-disease and miRNA-disease were modeled into bipartite networks and three methods, i.e. network based inference (NFI) algorithm [27], EF structure similarity-based inference model and disease phenotype similarity-based inference models were was used to generate an EF-miRNA-disease association model which is validated via 10-fold cross validation. The cases studies presented display impressive results. However, this work too, can predict associations between EF-miRNA-disease which are known in prior and does not predict novel associations [26].

Our models, in contrast to the previous works does not present miRNA-disease predictions, rather performs a maximum matching in a set of miRNAs and diseases to determine and prioritize diseases with highest cumulative impact. Hence, the resulting diseases, each of them have valid PubMed literature supporting it, and thereby accurate association with miRNAs. This gives the user complete confidence in the results, he/she is provided with. Further more, all other previous tools are prediction models, predicting *a* miRNA-disease edge/association. These models do not produce associations between set of miRNAs onto a set of diseases, thereby not exploring the overall dynamics of multi-level interaction of a miRNA-disease network.

Our model (in Chapter 4) which acts as an extension to the existing body of work in this field, works on a set of miRNAs and produces an output of a set of associated diseases, taking into account the impact and association of every miRNA in the set with every disease in the set.

In Chapter 5, the motivation was to infer miRNA-miRNA interaction networks based on miRNA-disease expression data and foldchange values. Computations efforts have been implemented to study and discover the disease-miRNA interaction network based on functional enrichment analysis[28], social network analysis method[29], similarity-based methods[7], diffusion-based method[30], Within and Between Score approach[31], integrating various genomic and phenotype data[32], using the model of Restricted Boltzmann machine[33] and based on support vector machine, among others. Similarly, co-regulated miRNA clusters across different types of diseases and the prioritized candidate miRNAs across multiple diseases have also been predicted. Such examples are- miRNA-miRNA synergistic network construction using functional modules and topology[34], by integrating multidimensional high-throughput data, through a progressive data refining approach, by grouping miRNAs based on their shared diseases, shared common targets and GO enrichment analysis of their predicted targets[35], by matching miRNA and mRNA expression profiles[36], through topological features in the deregulated miRNA target network[21] using cross-cancer differential co-expression network[37, 38], and by maximum weighted matching inference model and motif-based analysis method[39]. However, a 'systems model' is essential for identification of a core miRNA-miRNA co-regulatory signature or pattern across several cancers. Network theoretic algorithms such as biclique-based (graph theoretic) method[40], biclustering technique based on a bipartite graph method[38] among others have been deployed to discover and predict the patterns of miRNA regulatory models. Similarly, graph theoretical methods and network inference mod-

els were applied to analyze complex regulatory interactions and reconstruction of regulatory and other biological networks[41–44].

For Chapter 6, the motivation was to study the *social networking* aspect of the miRNAs. The study of information diffusion has always been an attractive area of research for the social and information-based research community. The concept of information diffusion in a network has been widely deployed in the field of social network theory to study spread of ideas, rumors and product adoption between the individuals in the network via the 'word of mouth' effect [45–47]. Although, this concept was initially applied in the field of sociology to study the various behavioral phenomena, like the spread of a new concept, it was also extended later to understand the spread of specific diseases [48]. However, understanding influence diffusion in a complex network like that of miRNAs' can be challenging. In this work, considering the fact that miRNAs of similar diseases tend to work together [20], we focus on the 'social nature' of miRNA and diseases and deploy an information diffusion model which is apt for such cases, wherein a miRNA's influence on its neighboring miRNAs is analyzed. Social influence can affect a range of behaviors in networks such as dissemination of information/influence, communication and in this case, even mutation. In both the LT and IC model, the nodes (i.e. the miRNAs) in the network can be in one of the two states - active or inactive. The activated nodes spread their influence by activating their neighboring inactive nodes based on a certain criteria or effect. Garnovetter et al. [49] proposed the LT model by applying a specific threshold in each of the nodes of the network. Herein, each node is activated only by its neighbor(s) depending upon the cumulative weight of the incoming edges to the node. The node becomes active when the cumulative sum of the weight of the incoming edges from an active neighboring node crosses its threshold value. Once activated, the node remains active and tries to activate its neighbor, thereby propagating its influence. On the

contrary, the IC model uses edge probability to determine the information diffusion. In this model, an active node has a single opportunity to activate its neighbors. The edge weights represent the activation probability or likelihood of information propagation in between two nodes. Hence, upon activation, an active neighbor may choose the neighbor with highest edge weight to activate.

The following chapters describes each methodology, results and findings in further depth and detail.

CHAPTER 3

MIREGULOME - MIRNA DATA AND ANALYTICS REPOSITORY

In *miRegulome*, we have incorporated all experimentally validated data for every module (upstream regulators, downstream targets, modulated pathways, regulated biological processes, and associated diseases) of a miRNA regulome from published literatures indexed in PubMed. *miRegulome v1.0* contains experimentally validated information for 803 miRNAs from 12 species, 113 chemicals, 187 upstream TF regulators, 3079 targets, and 160 diseases manually curated from 3417 PubMed indexed articles. Predicted 873 functions and 355 pathways are currently available in this database. The data repository comprises of

1. miRNAs and upstream chemical regulators : Each selected article having chemical-miRNA relationships is curated to capture the (i) chemical(s) (ii) miRNAs responding to the chemical(s), (iii) species of the miRNAs, (iv) expression of the miRNAs (up-/down-regulation) in responses to the chemical(s), (v) experimental conditions, (vi) techniques used to detect the expression levels of miRNAs, and (vii) the corresponding PubMed ID

2. Upstream TF regulators and downstream targets : experimentally validated upstream TF regulators and the downstream target genes/mRNAs of each miRNA that are having have upstream chemical regulators, are curated from the PubMed indexed literature and incorporated into the database.

3. Prioritized targets and miRNA functions of each miRNA using the Toppgene[50] suite of tools.

4. miRNA involved pathways

5. disease module : miRNA-disease relationships along with regulation of the miRNA (up- and down-regulation) in the disease condition were curated from PubMed listed published literature for those miRNAs that respond to chemical stimulus and were incorporated

3.1 Analyses tools

miRegulome is empowered with four unique tools to provide meaningful associations among chemical-disease, miRNA-disease, gene-disease, and disease-chemical-miRNA along with affected BPs based on user specific datasets. The results of these analyses correspond with a bipartite modeling approach which we developed in Chapter 4 to explore the associations among miRNAs and diseases available in the database. In *miRegulome*, each association whether its chemical-miRNA, miRNA-disease, gene-miRNA etc. is manually curated from PubMed indexed literature and each of such these relationships are tagged with specific PubMed ID from where the data are taken. These tools mine the database and give relevant associations to the user by querying the miRegulome database and counting the associations (direct or indirect) between entries. The output is returned as ranked association counts with Z-score statistical analysis rather than statistical enrichment measures. We calculated the Z-scores using the formula: $Z - score = (X - \mu)/\rho$, where X is the association count of the particular association i.e. the number of PMIDs citing the association, μ is the mean of the association counts for the entire association type, and ρ is the standard deviation. Hence, a positive $Z - score$ indicates that the count of the association is higher than the average of such associations, and a negative value indicates that it is below the average mean. A value of 0 would mean, it is equal to the average.

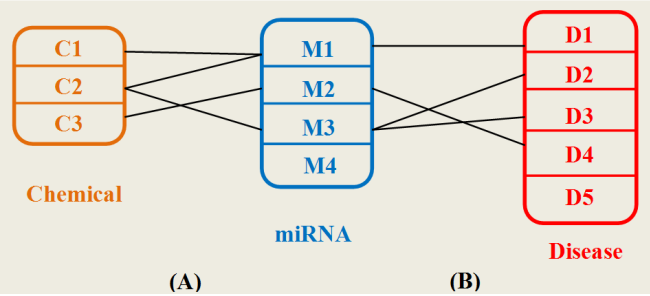
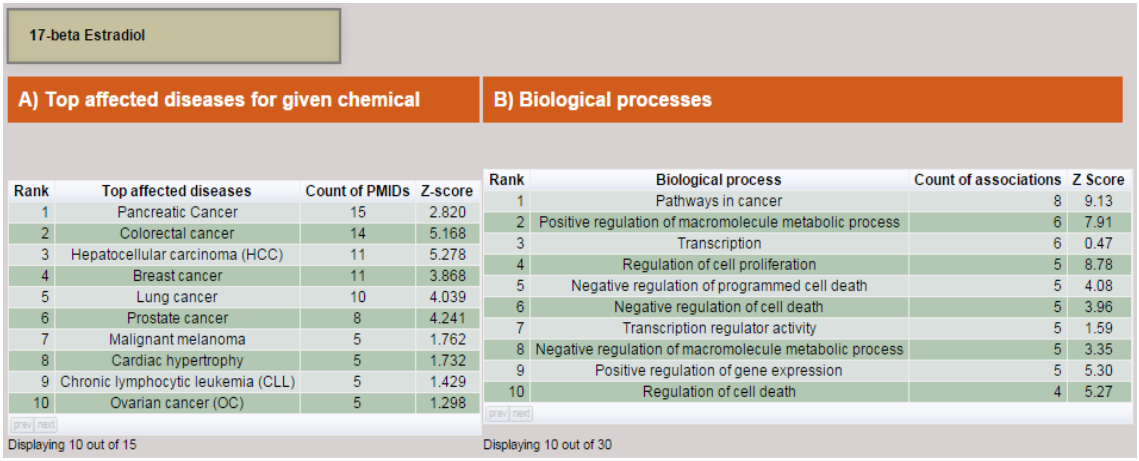


Fig. 2.: Chemical-miRNA-disease tool

Chemical-disease analysis

This analysis tool allows the user to explore the associations between a chemical to a disease via miRNAs. When a user selects a particular chemical, the tool retrieves all the miRNAs associated with the chemical. Thereafter, the tool retrieves all the diseases in which the miRNAs are associated. For example, if a user selects chemical $C2$ (see Figure 2) in the Chemical-Disease analysis tool, miRNAs M1 and M3 are retrieved and subsequently, their associated diseases $D1$, $D2$ and $D3$ are retrieved. Finally, the tool ranks the diseases in which these miRNAs (which are associated with the chemical) are associated, counting the PubMed IDs. The tool then displays the disease names, their association counts and their respective Z-scores for the counts (See Figure 3). Using similar methodology, the BPs which are associated with the miRNAs are displayed according to the count of their associations as recorded in the database. It does not assert a direct link between the chemical to a disease or to the BPs via the miRNA, rather allows the user to explore and test their hypothesis for indirect associations between the chemical and the disease via the miRNA.



17-beta Estradiol

A) Top affected diseases for given chemical

Rank	Top affected diseases	Count of PMIDs	Z-score
1	Pancreatic Cancer	15	2.820
2	Colorectal cancer	14	5.168
3	Hepatocellular carcinoma (HCC)	11	5.278
4	Breast cancer	11	3.868
5	Lung cancer	10	4.039
6	Prostate cancer	8	4.241
7	Malignant melanoma	5	1.762
8	Cardiac hypertrophy	5	1.732
9	Chronic lymphocytic leukemia (CLL)	5	1.429
10	Ovarian cancer (OC)	5	1.298

prev next

Displaying 10 out of 15

B) Biological processes

Rank	Biological process	Count of associations	Z Score
1	Pathways in cancer	8	9.13
2	Positive regulation of macromolecule metabolic process	6	7.91
3	Transcription	6	0.47
4	Regulation of cell proliferation	5	8.78
5	Negative regulation of programmed cell death	5	4.08
6	Negative regulation of cell death	5	3.96
7	Transcription regulator activity	5	1.59
8	Negative regulation of macromolecule metabolic process	5	3.35
9	Positive regulation of gene expression	5	5.30
10	Regulation of cell death	4	5.27

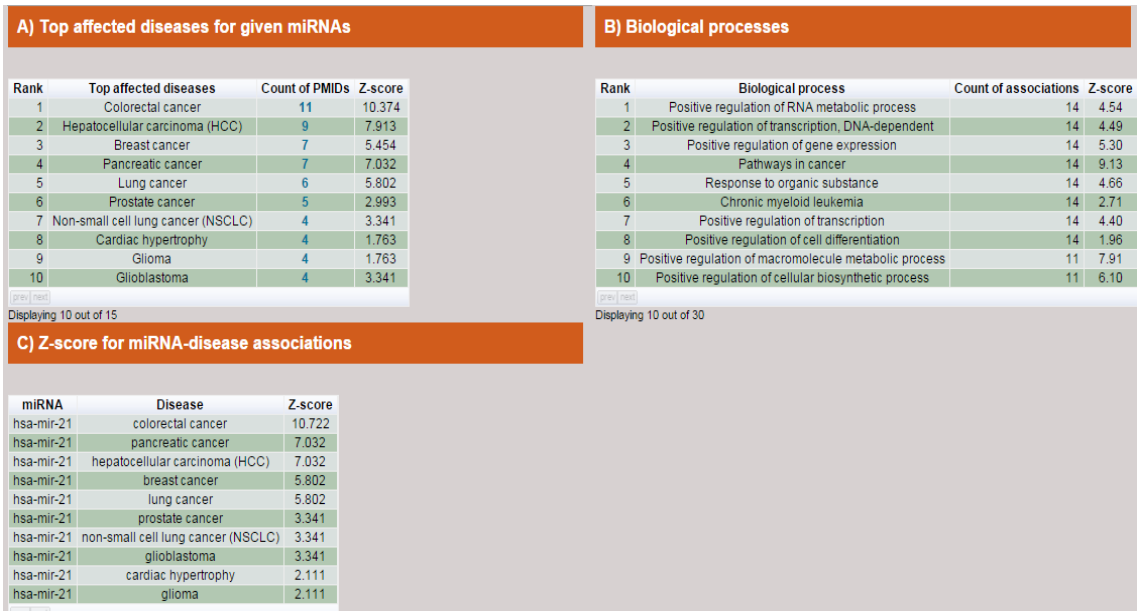
prev next

Displaying 10 out of 30

Fig. 3.: Chemical-miRNA-disease tool results

miRNA-disease analysis

In this analysis, when an input of one or more miRNAs is provided, the tool provides three tables for the user to get a comprehensive understanding of their results (See Figure 4). The tool searches for all diseases associated with the provided miRNA (s) and the distinct miRNA-disease associations (based on PubMed IDs). Following which, it ranks the diseases based on their number of recorded (PubMed IDs) associations and displays them. The user can click on the $CountofPMIDs$ and see the unique PubMed IDs supporting the results. The tool also displays the Z-scores for each disease along with its rank. Z-score here is a standardized score for the count of each disease, indicating the resultant disease's location in a distribution of other diseases, in relation to the mean and standard deviation of miRNA-disease counts. The Z-score of the disease tells the user, how many standard deviations it is from the mean distribution of all miRNA-disease count distribution present in the database. To further understand the relative impact of each miRNA (entered by the user) to the disease, we display each miRNA-disease edge with a Z-score. This value gives the user the, individual miRNA-disease strength of association. It also displays which among



A) Top affected diseases for given miRNAs

Rank	Top affected diseases	Count of PMIDs	Z-score
1	Colorectal cancer	11	10.374
2	Hepatocellular carcinoma (HCC)	9	7.913
3	Breast cancer	7	5.454
4	Pancreatic cancer	7	7.032
5	Lung cancer	6	5.802
6	Prostate cancer	5	2.993
7	Non-small cell lung cancer (NSCLC)	4	3.341
8	Cardiac hypertrophy	4	1.763
9	Glioma	4	1.763
10	Glioblastoma	4	3.341

Displaying 10 out of 15

B) Biological processes

Rank	Biological process	Count of associations	Z-score
1	Positive regulation of RNA metabolic process	14	4.54
2	Positive regulation of transcription, DNA-dependent	14	4.49
3	Positive regulation of gene expression	14	5.30
4	Pathways in cancer	14	9.13
5	Response to organic substance	14	4.66
6	Chronic myeloid leukemia	14	2.71
7	Positive regulation of transcription	14	4.40
8	Positive regulation of cell differentiation	14	1.96
9	Positive regulation of macromolecule metabolic process	11	7.91
10	Positive regulation of cellular biosynthetic process	11	6.10

Displaying 10 out of 30

C) Z-score for miRNA-disease associations

miRNA	Disease	Z-score
hsa-mir-21	colorectal cancer	10.722
hsa-mir-21	pancreatic cancer	7.032
hsa-mir-21	hepatocellular carcinoma (HCC)	7.032
hsa-mir-21	breast cancer	5.802
hsa-mir-21	lung cancer	5.802
hsa-mir-21	prostate cancer	3.341
hsa-mir-21	non-small cell lung cancer (NSCLC)	3.341
hsa-mir-21	glioblastoma	3.341
hsa-mir-21	cardiac hypertrophy	2.111
hsa-mir-21	glioma	2.111

Fig. 4.: miRNA-disease tool results

the input miRNAs has the highest/lowest impact on the disease, thereby giving a more in-depth insight into the results. Moreover, the 'Count of PMIDs' gives the cumulative count of PubMed IDs citing the associations of the input miRNAs with the disease.

Gene-disease analysis

When a list of genes is entered by the user in the input field, the tool searches for miRNAs associated with the set of genes and counts the number of gene-miRNA associations (i.e. PubMed IDs) recorded in the database. Thereafter, the tool searches and counts the existing relationships (i.e. PubMed IDs) between the observed miRNAs and diseases. Following which, the tool ranks the diseases based on their count of PubMed entries. Similarly, the tool also displays the list of BPs which are associated with the specified set of genes via miRNAs, and ranks them following the same principle of the miRNA-disease analysis tool. The tool does not assert a relation-

ship between the entered genes and diseases but highlights the top diseases indirectly associated with the genes entered, via the miRNAs.

Disease-chemical/miRNA analysis

This tool works in the opposite way of the Chemicaldisease analysis tool. It takes disease(s) as an input and searches the repository for specific miRNA(s)-disease associations. Thereafter, it retrieves the chemicals associated with the miRNAs. The tool displays these associations and ranks them based on the number of occurrences in the database (i.e. PubMed IDs). This gives the user an insight into possible role of chemicals in regulating miRNAs which are deregulated in the input disease(s).

miRegulome aims to provide the complete regulome of any miRNA listed in this database as derived from published literature. The current version of miRegulome v1.0 provides the complete regulome for chemically responsive 803 miRNAs from 12 species. *miRegulome* can be accessed online at http://bnet.egr.vcu.edu/miRegulome and is free for academic research. jQuery, JavaScript, and HTML has been used to design the web-user interface. PHP has been used for server-side scripting support. Google visualization library has been used to represent the data in an interactive form. MySQL v.5.1.63 is used as database to store regulome information. It runs on Apache 2.2.17 on Ubuntu 12.04. The database and its integrated tools can be best used using the latest versions of Google Chrome and Mozilla Firefox browsers.

This work was published in Barh et al., 2015.[51] A future work in alignment with this motivation was also published in Nalluri et al., 2016.[52]

CHAPTER 4

MAXIMUM WEIGHTED MATCHING MODEL

Identification of miRNA-disease associations through experimental laboratory methods are time consuming and expensive [4]. Hence, a large interest has been devoted towards finding important underlying associations through various computational models.

A network of miRNAs and diseases underlain with TFs and target genes is a very dense network and thereby poses a very complex networking scenario. Complex networks offer a unique perspective to explore relationships among homogeneous and heterogeneous entities. These entities can be biological molecules, diseases, genes etc. Hence, graph theoretic concept is very apt to model and mine important miRNA-disease associations. In our research, almost all the observed miRNA-disease networks, such as *miRegulome*, *mir2Disease* [9], miRNA-disease association network (MDAN) [19] and Human MicroRNA Disease Database (HMDD) [20] are scale-free; meaning few nodes i.e miRNAs have the highest impact on other nodes, thereby acting as hubs. Hence, a miRNA-disease network follows the topological characteristics of scale-free networks. For e.g. Figure 5 shows a scale free network of miRNA-disease association network of HMDD. Further details about the topological metrics of the scale-free nature of these miRNA-disease networks are elaborated in the Section *Motif-Based Analysis*.

Using the graph theoretical network model, in this work we aim to find the most impacted diseases upon action/altercation of specified miRNAs. Here, we present a model that determines a prioritized set of diseases which are most definitely in-

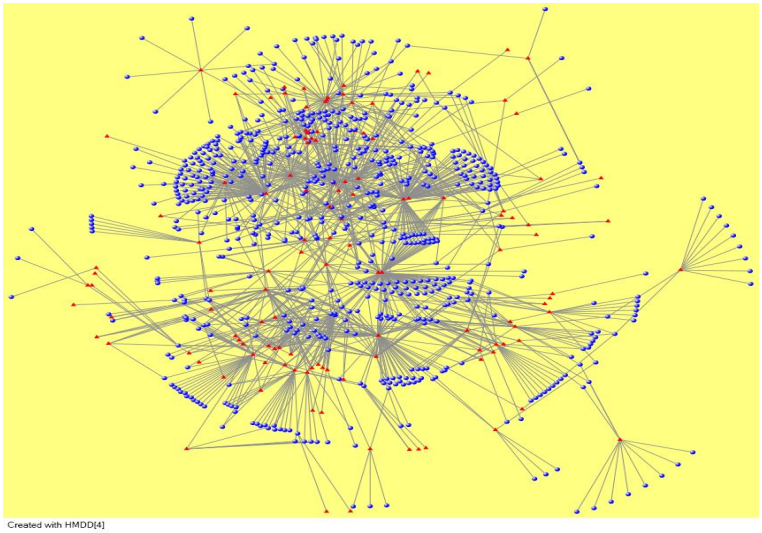


Fig. 5.: Network of miRNA disease associations in HMDD. Blue circles represent miRNAs and red triangles represent diseases

fluenced upon the cumulative action/altercation of specified miRNAs. These associations are determined by a pipeline process of applying the maximum-weighted-maximum-matching algorithm to the network model in Section 4.1, calculating cumulative weights per disease in Section 4.1.1, and applying the disease ranking scheme in Section 4.1.2. A preliminary version of this work was presented earlier[53]. Furthermore, none of the previous work have presented any work on the motif-based analysis of miRNA-disease networks. In this chapter, we analyze the topological features of several miRNA-disease networks, especially the motifs in these networks and also the cumulative impact of a set of miRNAs onto a set of diseases. The visualization of these results and their topological perspective is elaborated in the Section 4.2.3.

4.1 Maximum Weighted Matching Inference model

Single or multiple miRNA(s) is/are up- or down- regulated in one or a set of disease(s). The instances of up and down-regulations between a miRNA and disease,

signify the strength of association between the pair. The interactions of miRNAs and diseases can be mapped as a complex network such that miRNAs and diseases are nodes in the network [Figure 5]. This mapping is critical to explore the associations and depends heavily on the type of interactions. A graph theoretical concept such as bipartite graph [54] can be used to model this problem. In this work, we have modeled the miRNA-disease interaction as a bipartite graph which is shown in Figure 6-A.

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 [54]. In our model, miRNAs and diseases have been categorized as two disjoint sets and an edge represents an association between them. The data consisting of miRNAs and diseases has been used from *miRegulome*. Herein, the edges are weighted i.e. the number of publications citing up/down regulations between a miRNA-disease pair. For e.g., in Figure 6-A the edge weight of 20 between $m1$ and $d1$ represents the number of PubMed IDs citing miRNA $m1$ regulating disease $d1$. Hence, the weight of the edge represents the strength of the association between the miRNA and disease. Based on this data, we derive a weighted network consisting of miRNA-disease interactions.

Maximum weighted matching (MWM): 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 with largest possible set of edges. A maximum weighted matching is a maximum matching such that the sum of the weights of the edges is maximum. This is explained below.

Consider a miRNA-disease interaction network as in Figure 6, where $m1$ to $m4$ are miRNAs and $d1$ to $d7$ are diseases. The weight on the edge represents the strength of the association between the miRNA-disease pair, in terms of the number of pub-

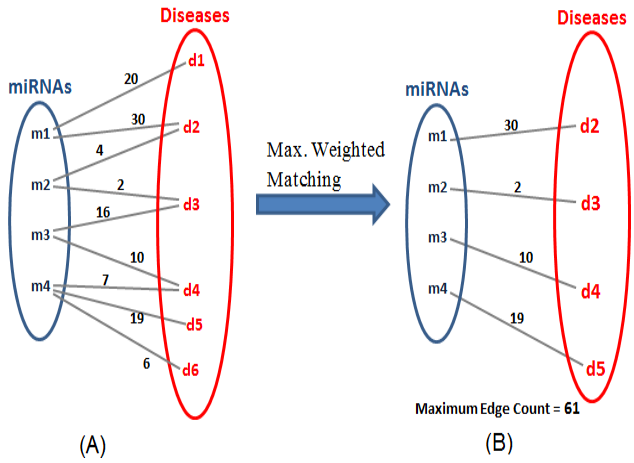


Fig. 6.: Maximum weighted matching

lications citing up-regulating and down-regulating a disease. For e.g. in Figure 6, the edge $m1$-$d2$ has a weight of 30, which indicates there are 30 publications (i.e. PubMed IDs) in the curated literature of *miRegulome* which cite the miRNA $m1$ either up-regulating or down-regulating disease $d2$. As Figure 6-(B), shows after the application of the MWM algorithm, the resultant sum of edges is the maximum score, which implies that there is no possible *combination* of m-d pairs in the network, whose cumulative sum is higher than the result. Hence, the MWM helps in determining the strongest miRNA-disease pairs combination among 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 to present a concise list of diseases with highest confidence of being influenced and not to determine specific miRNA-disease associations; rather an association between a *set* of miRNAs onto a *set* of diseases. Models of such association that calculate the cumulative impact of a set miRNAs onto a set of diseases are not many. This is important 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.

In our premise, since we are exploring the associations among a set of miRNAs onto a set of diseases, it is important to note that many diseases might be associated, but not all diseases might be significantly relevant to the set of inputted miRNAs. Hence, we have to consider each miRNAs sphere of influence onto diseases, as well as its relevance to other miRNAs sphere of influence. Herein, the MWM algorithm addresses the issue, by choosing the optimum set of associations (with highest cumulative sum) that the set of miRNAs present. This algorithm takes into account each miRNAs sphere of influence and its strength of influence/association and thereby calculates a set of edges, in consideration with set of miRNAs such that the resultant cumulative influence of the set of miRNAs onto the set of diseases is highest. In other words, just because a certain miRNA-disease edge has not been selected, it does not imply, it is not considered. What it implies is that, it is not important when the entire set is considered. Also, the goal is to produce a concise list and not an entire set of associated diseases. This constraint does well to generate a set which is both representative of every miRNAs sphere of influence as well as determining the highest impacted diseases.

In any given miRNA-disease network, the solution to the MWM algorithm in a given $G(V, E)$ can be solved as an optimization problem.

Optimization problem formulation: Objective: To achieve the maximum sum of weighted edges between miRNA and diseases, subject to constraints that no vertices share the same edge. This helps us in getting the most prominent collection of pairs such that, their cumulative sum is the maximum among all possible combinations.

Variables: Let $X_{i,j}$ be an edge between a miRNA and disease, $Weight_{i,j}$ be the edge weight between the miRNA-disease pair, m and d be the set of miRNAs and

diseases respectively.

Algebraic formulation:

$$Maximize \sum_{i,j} Weight_{i,j} * X_{i,j}$$

$$s.t. \sum_{j} X_{i,j} \leq 1 (j = 1, 2, ..., m)$$

$$\sum_{i} X_{i,j} \leq 1 (i = 1, 2, ..., d)$$

In the above formulation, we are maximizing the cumulative sum of the edges, with the constraints that no miRNA or disease should be repeated. These constraints help in reducing the repetition of common diseases associated with different miRNAs; since miRNAs tend to regulate about 50 to 100 or more diseases based on data in the human microRNA disease database (HMDD)[20] and *miRegulome*. This is important keeping in view that the goal is to present a breadth of diseases *within* the concise list, bearing on the fact that miRNAs tend to work closely in sets.

The above MWM optimization formulation is a linear programming problem and geometrically, its a convex function. The resulting feasible region of solutions is a polyhedron. This linear programming equation is solved using the linear program (LP) solver GLPSOL which uses the simplex method [55].

4.1.1 Prioritization of disease candidates

Since many miRNAs are connected to a single disease they have a cumulative influence on it. For e.g. in Figure 7, disease $d2$ is influenced by miRNAs $m1$ and $m2$. Similarly, diseases $d3$ and $d4$ are influenced by more than one miRNA. In real scenarios, diseases are regulated by multiple miRNAs. Hence, it is vitally important that we consider the cumulative impact of all the active miRNAs on its associated

diseases. In this model after the miRNAs-disease network is created based on user input of active miRNAs, we calculate the cumulative impact of all of them on each connected disease. Figure 7, shows the influence on each diseases numerically. This helps in understanding in many ways, how a disease can be influenced by multiple miRNAs, which is not considered in the MWM model. The MWM model, as shown in Figure 6, selects the top impacted diseases. Each diseases' impact can be calculated by adding the weights of every active miRNA and the particular disease, as shown in Figure 7.

This approach gives a ranked list of diseases.

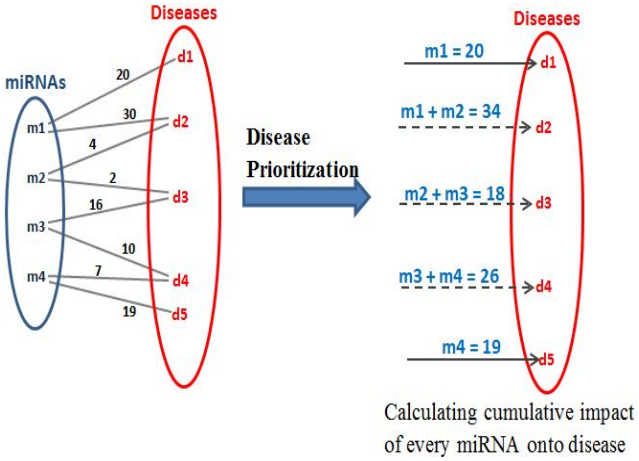


Fig. 7.: Cumulative impact of each miRNA

4.1.2 Disease ranking scheme

Although, the application of MWM algorithm gives the most prominent miRNA-disease associations, it has a limitation. Because of the constraint that no two edges can share a common vertex, a strongly associated miRNA-disease pair can get ignored in the MWM selection process. For example, consider miRNAs $m2$ and $m3$ in Figure 6; for miRNA $m2$ and miRNA $m3$, the $m3 - d3$ pair weight is 16 and $m3 - d4$ pair

weight is 10. However, in the resultant matching only $m3 - d4$ pair is selected (see Figure 6), because addition of *this* edge provides the highest cumulative sum when all possible resultant combinations are considered. The pairs $m3 - d4$ and $m2 - d3$ are selected in the matching but their pair weights are 10 and 2 respectively, which is less than the non-selected pair, $m3 - d3$. In order to overcome this limitation, a disease ranking scheme has been adopted.

Here, diseases are ranked as per their highest cumulative impact from miRNAs (see Figure 8-C) as explained in Section 4.1.1. This set of ranked diseases is compared with the set of diseases obtained after the MWM algorithm (See Figure 8-B). The rank of the disease in the MWM set which is least ranked is noted. If there are other diseases which have a higher rank than the least-ranked disease ***and*** are not included in the MWM set; those are added to the final output set of diseases (see Figure 8-D). This method makes sure that a disease which is highly influenced is not missing after the MWM algorithm is applied. MWM algorithm helps in giving a definite and concise set of affected diseases. Prioritizing of diseases ranks them as per their impact. Disease ranking scheme enhances the result set by overcoming the limitation of the MWM algorithm, and adding higher ranked diseases in the final resultant set of diseases.

When the miRNAs are entered by the user, an automated script performs the following functions:

1. Runs the database procedure gathering the relevant literature pertaining to the set of miRNAs

2. Generates the cumulative impact of miRNA onto each disease in a ranked manner (Figure 8-C)

3. Creates a network model of the miRNA-disease associations in GMPL (Figure

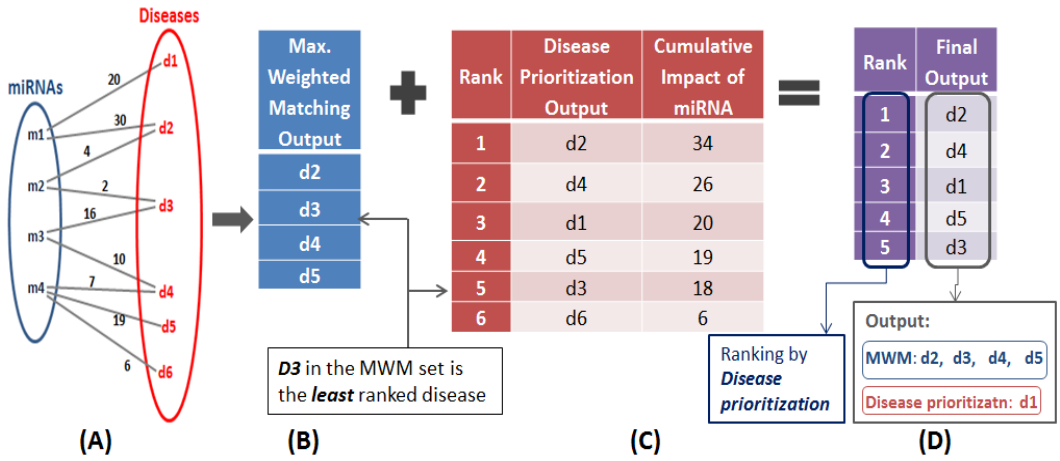


Fig. 8.: Cumulative impact of each miRNA

8-A)

4. Runs the MWM optimization script which operates on the created network model and generates the optimum set of associations (Figure 8-B)

5. Observes the disease in the results of MWM (Figure 8-B) and identifies the least ranked among them. Thereafter, it checks for diseases with higher cumulative count than the least ranked disease, in the result set of (2), i.e Figure 8-C. If there are any diseases with higher cumulative impact and not included in the MWM set (4), they are added to the resultant set (Figure 8-D). For e.g. in Figure 8, the set of diseases through MWM were $\{d2,\ d3,\ d4,\ d5\}$ and the set of diseases through 'Disease ranking' were $\{d2,\ d4,\ d1,\ d5,\ d3\}$. Disease *d1* had higher cumulative impact compared to the least ranked disease *d3* in the MWM set and hence it was added to the final resultant section. Therefore, the final resultant set of diseases is $\{d2,\ d4,\ d1,\ d5,\ d3\}$

This model has been used on the data from *miRegulome*, HMDD, and *miR2Disease* [9] databases. Table 1 presents some of the results. PubMed IDs are provided for

further reference.

This approach stands in contrast with many of the previous approaches mentioned in the *Literature* section. Firstly, most of the previous works, for e.g. [19],[4],[5] and [6] are *prediction* based results and present *'1-to-1'* miRNA-disease association. In contrast, our work explores the associations between the *set* of miRNAs onto *set* of diseases and presents results which are *known* associations, validated by PubMed ids and not predicted. Secondly, our starting premise and motivation for this work, unlike the previous works, is to explore the collaborative working of the sets of miRNAs and diseases. There are not many tools, which determine a set of diseases based on the user's input set of miRNAs to which we can compare. Thirdly, previous works present a list of associations between miRNA and diseases which are static in nature, and predict *new* associations which are valid with certain confidence score. However, the approach and results in our work are dynamic; meaning the results will change every time a new set of miRNAs are entered. The results are generated at the instant - by *sending a query* to gather the relevant literature, *generating* the network model, *optimizing* the objective of the network model, *calculating* the cumulative impact on diseases and *producing* the set of diseases. As more and more new associations are added to the databases, the results would only change for better. The results are not *new* predictions rather set of *known* diseases, determined and prioritized to the set of input miRNAs. Owing to the aforementioned reasons, there could not be a reasonable and fair comparison done with previously established, benchmarked prediction-based datasets used in [19], [6], [7] and [23] which are static, 1-on-1 miRNA-disease predictions.

This work was published in Nalluri et al.,2013.[53]

4.2 Motif based analysis

The topological features of miRNA-disease network could provide valuable insights into the nature of collaboration of miRNAs and diseases, since miRNAs emerge to work in groups [20]. It has been observed that motifs are the fundamental building blocks in biological networks [56], since they are frequently occurring substructures. These substructures can be of sizes 2 or above. Hence, we studied the topological features of this network, namely motifs. We performed a motif-based analysis of a miRNA-disease interaction network, and the disease-disease interaction network. *mfinder* [57] and *fanmod*[58] software are used to determine the most significant motifs in the considered miRNA-disease networks. Motifs generated by *mfinder* are identified in green color and motifs generated by *fanmod* are identified in orange color. Apart from the networks derived from *miRegulome*, these motif-based analyses were also performed in *miR2Disease*[9] network and also the HMDD [20] database.

4.2.1 mirna-disease network

The miRNA-disease associations obtained from *miRegulome* contained 468 nodes and 2998 edges which is a sparse network with a density value of 0.0273. The degree distribution of miRNA-disease network (see Figure 9) follows power-law property of scale-free networks, i.e. their degree distribution follows the property of $P(k) \sim k^{-\gamma}$ [59]. Earlier research on scale-free networks showed that such networks are modular. While bipartite graph analyses identify diseases that are most influenced by miRNAs using empirical evidence, motif analyses offers an additional perspective by introducing structural insight to the miRNA-disease networks. The following 3 node (see Figure 10) and 4 node motifs (see Figure 11) were found to be significant. The 3 node motif implies there a miRNA regulating at least two diseases and at least two

miRNAs regulating a single disease. It also corroborates the finding that, when a single disease is being regulated by a single miRNA, that same miRNA is regulating one another disease - thus implying a non-direct way (i.e. via a miRNA) of a disease affecting another disease. Hence, if two miRNAs are regulating a single disease, it can be deduced that either the miRNAs are working against each other or in agreement with each other in regulating that particular disease. A significant presence of this motif implies, there are multiple connections of this sort among a diverse set of miRNAs and diseases, which pose a complex networking scenario. A significant amount of 4 node motifs in this network emphasize the earlier observation made above; in that a single miRNA is regulating three diseases, three miRNAs regulating a single disease and two miRNAs regulating a two diseases. This provides a glimpse into the intricate networking of miRNAs and diseases. These results are further corroborated by the findings of MDAN [19], that 64.96% of diseases were *at least* associated with two miRNAs and about 70% of the miRNAs were associated with two or more diseases. In the 3-node motif and the 4-node motif, the nodes could represent either a miRNA or a disease. However, the edge will always represent an association between a miRNA and a disease. Hence, if a certain node is assumed to be a miRNA, the node lined to it is a disease and vice-versa.

In *DISMIRA* — a tool developed based on the approach presented in this paper — upon the input of miRNAs, the top diseases are displayed which participate in maximum number of motifs in the network of entered miRNAs and diseases. Visualization presents an insightful display of the motif structures, thereby providing the research community with a graphical understanding of the nature of association between the miRNAs and diseases.

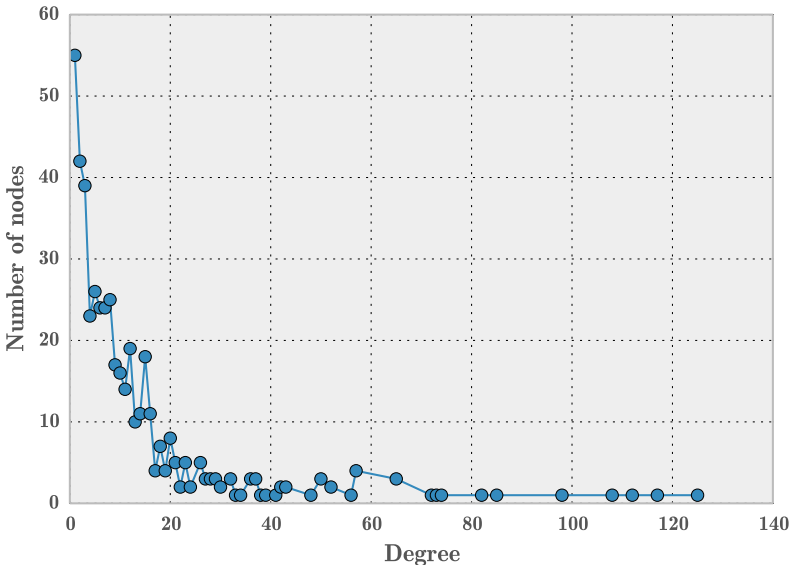


Fig. 9.: Degree distribution of miRNA-disease network

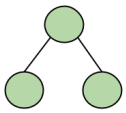


Fig. 10.: 3 node motif in miRNA-disease network

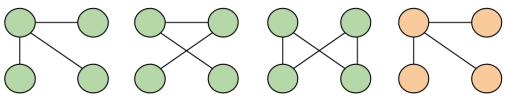


Fig. 11.: 4 node motifs in miRNA-disease network

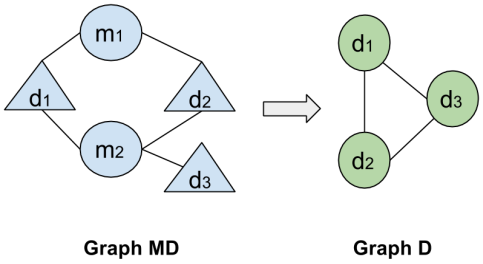


Fig. 12.: Example showing the graph transformation from miRNA-disease network to disease-disease network

4.2.2 Disease-disease network

In order to understand the associations and pattern between the diseases, an exclusive disease-disease network was derived as a projection off the miRNA-disease network. Consider the miRNA-disease network to be graph MD and the disease-disease network to be graph D. An edge between two diseases exists in D if both these diseases are influenced by the same miRNA. This graph transformation is demonstrated in Figure 12.

The resulting network has 132 nodes and 3357 edges. To determine the structural properties of disease-disease network, its degree distribution is plotted in Figure 13. Observably, the distribution does not seem to follow power-law distribution which usually indicates scale-free nature of a network. Upon analyses, the same motifs (see Figure 10 and Figure 11) which were observed in miRNA-disease network, were found to be significant in this network. This observation supports the notion that diseases tend to work in tandem with other diseases and in our case via a miRNA passage, they influence each other.

These same motifs were observed to be significant in the mirna-disease network of the human microRNA disease database (HMDD) [20] and the *mir2Disease* database

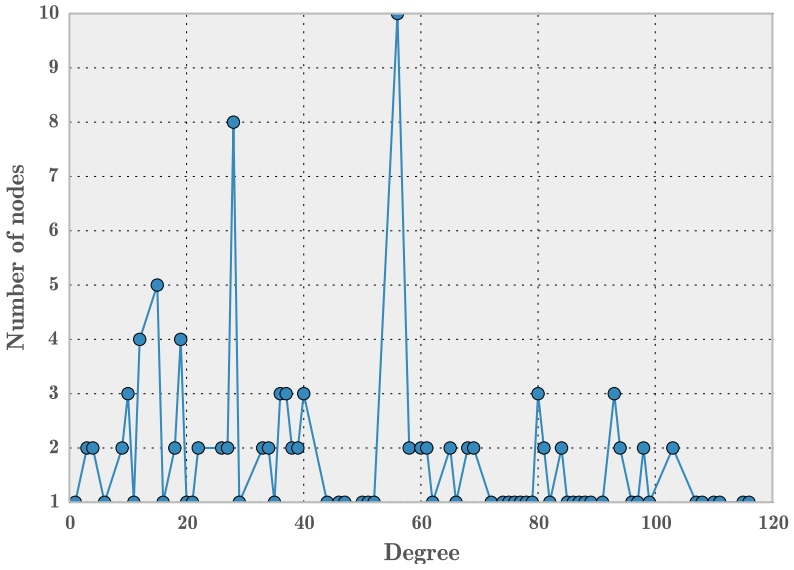


Fig. 13.: Degree distribution for disease-disease associations

[9], hence strengthening the case for these motifs to be vitally important in the available miRNA-disease networks. The HMDD network consisted of 961 nodes (of miRNAs and diseases) and 6448 edges, while the *mir2Disease* database consisted of 309 nodes (miRNAs and diseases) and 637 edges.

mfinder and *fanmod*, both generate random networks in their process of motif identification. *mfinder* uses 100 random networks and *fanmod* uses 1000 random networks. During this randomizing, 4-node or 3-node sub-graphs are generated among which, the identified motifs have been found to be significant. Figure 14 is an excerpt of the result summary for the significance of 4 node motif in the miRNA-disease network of *miRegulome* by *mfinder*. The explanation has been taken from the manual guide of *mfinder*.

Figure 14 explains the number of occurrences of the 4-node motif in the network, the criteria taken for a motif to be significant, its Z-score, uniqueness and number of random networks generated.

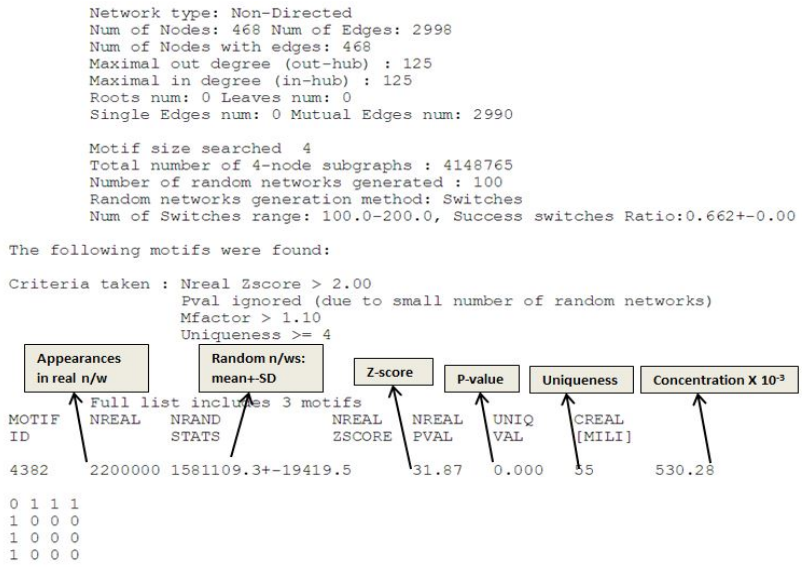


Fig. 14.: Summary of results explaining motif identification/significance

We have incorporated motif-based analysis feature in the tool *DISMIRA*. Upon the input of miRNAs, the tool will display the diseases which have the highest sharing of motif structures with other miRNAs/diseases.

Table 2 shows the diseases and respective motif participation counts for an example input set of miRNAs. Malignant melanoma, Epithelial ovarian cancer (EOC), Breast cancer and Lung cancer are found in thirty seven square motifs.

4.2.3 Network Visualization

The network of miRNAs and diseases can be easily observed in this interactive visualization feature of *DISMIRA*. This insightful perspective into the miRNA-disease associations helps the user in the understanding of the networking of miRNAs and their associated diseases, and also interpreting the the associations among miRNAs and diseases. Maximum weighted matching algorithm and the motif-based analyses are deployed into *DISMIRA* and their results are presented using the interactive visualization. The user can input a set of miRNAs and select either the maximum

weighted matching algorithm or the motif-based approach to identify significantly associated diseases, and see their corresponding regulations and PubMed IDs. Upon submitting the input query of miRNAs, the resultant diseases are displayed visually. miRNAs are represented by blue nodes, resultant diseases i.e. the top affected diseases from both the approaches are represented by orange nodes and other associated diseases to the miRNAs are represented by the green nodes (see Figure 15). The resulting miRNA-disease associations are represented using a network visualization in a force-directed layout, meaning placement of miRNAs and diseases are in the most aesthetic way and there is minimal crossing over of edges. This layout makes the understanding of the network very intuitive. Once the results appear, users can zoom-in and zoom-out of the graph for granular level of details such as edge associations, nearby entities and their respective associations etc. Edges between the nodes are disabled by default and are shown upon selecting a specific node. This helps in user-driven network discovery. Upon clicking a miRNA or a disease, its edges are highlighted giving the user, the immediate reach of the entity. Multiple node selections are available to identify nodes of common interest. Interacting with this network visualization of miRNAs and diseases provides helpful insights which are not collected otherwise, such as — the shortest path from a miRNA/disease to another miRNA/disease, the k or $k+1$ closest neighbors of a miRNA/diseases and a global perspective of a miRNA or disease's topological placement in the larger picture of this network.

There are no miRNA-disease interactive visualization tools available for free, as far as we know, at the time of this publication. The visualization tool in this work, generates a user specified network of miRNAs-diseases and allows the user to discover the network with the progression of clicks. The width of the edges i.e. thick and thin, intuitively convey the strength of association between the miRNA-diseases.

Furthermore, users can research the top impacted diseases and their subsequent up-/down regulation by miRNAs on clicking the disease details and searching them in PubMed literature (see Figure 16). Moreover, the results are displayed intuitively in a 3-way approach; after receiving the list of diseases in the output, the user can click on a certain disease and know *why* the disease is significant (based on the PubMed id count), *where* is it relevant (based on its topological position in the larger picture of miRNA-disease network) and *how* it is impacted (by seeing each miRNA's impact and subsequent regulation towards it). This would assist in thorough investigation. To aid in further detail and completeness for the user, once the network is displayed, all miRNA-disease associations along with their PubMed IDs are provided for download in CSV format. Users use this CSV file in other visualization softwares of their choice too. The visualization is developed using Django framework [60], Python [61] (networkX [62]), JavaScript [63], d3js library [64], bootstrap and HTML with the support of MySQL for back-end database. A snapshot of the visualization is presented in Figure 15. This tool can be accessed for free at: *http://bnet.egr.vcu.edu:8080/dismira.*

4.2.4 Case study and utility

Consider the input of miRNAs, hsa-mir-125a, hsa-mir-34a, hsa-mir-21 to *DISMIRA*. Upon choosing the maximum weighted matching (MWM) based model the most impacted diseases are: colorectal cancer, hepatocellular carcinoma (HCC), pancreatic cancer and breast cancer. However, users can select individual miRNAs and observe the association onto other diseases along with the strength of the association. Thicker edges represent high count of PubMed literature supporting the association and regulation (see Figure 15).

Moreover, upon clicking a certain disease, in our example, say 'colorectal cancer' - its subsequent regulation details, PMIDs can be retrieved. In this case, by studying

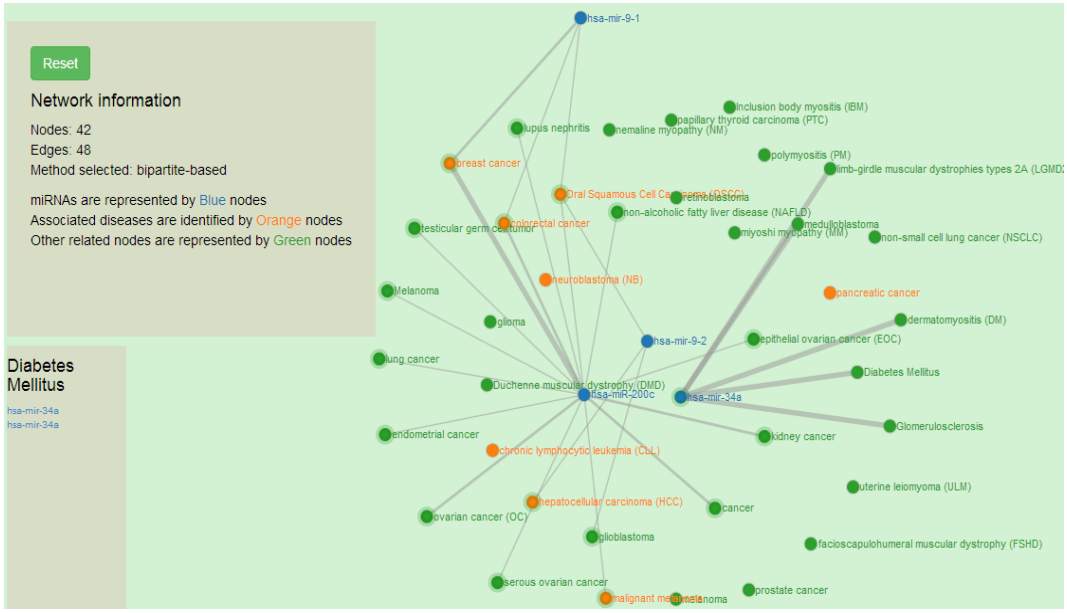


Fig. 15.: Visualization of results of maximum weighted matching algorithm

the results further, it can be noted that mir-21 is strongly up-regulated during this disease, whereas mir-125a and mir-34a are being down-regulated (see Figure 16)

However, in case of 'pancreatic cancer', mir-21 and mir-125a are both being up-regulated and mir-34a is being down-regulated. The scenarios of multiple miRNAs working together and against each other towards their regulation during a certain disease can be easily observed and studied. Upon selecting the motif-based approach for the same miRNAs, the top diseases are: hepatocellular carcinoma (HCC), colorectal cancer, prostrate cancer and pancreatic cancer with their motif counts of 470, 446, 431 and 317 respectively. These diseases are occurring in most motif structures of 3-node and 4-node. As shown in Figure 17, it is intuitive that these diseases would be in most of the motifs due to their topological placement in the network, i.e. prostate cancer, pancreatic cancer and hepatocellular carcinoma are the bordering diseases of three miRNAs' range. Hence, they are most participative in the interaction of several diseases i.e. 3-node and 4-node motifs.

colorectal cancer

miRNA	Pubmed ID	Regulation
hsa-mir-34a	19843336	Down-regulated
hsa-mir-34a	17875987	Up-regulated
hsa-mir-21	16461460	Up-regulated
hsa-mir-21	19843336	Up-regulated
hsa-mir-21	19509156	Up-regulated
hsa-mir-21	19826040	Up-regulated
hsa-mir-21	16609010	Up-regulated
hsa-mir-21	16854228	Up-regulated
hsa-mir-21	17968323	Up-regulated
hsa-mir-21	18196926	Up-regulated
hsa-mir-21	19737943	Up-regulated
hsa-mir-21	18230780	Up-regulated
hsa-mir-125a	19843336	Down-regulated
hsa-mir-125a	16609010	Down-regulated

Search all in Pubmed

Fig. 16.: Regulation details of colorectal cancer

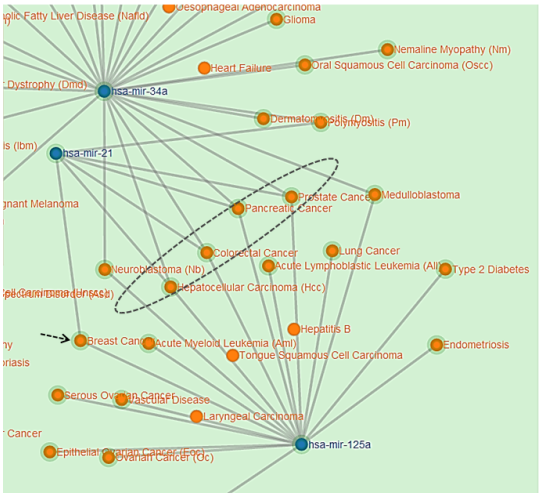


Fig. 17.: Motif participation of diseases

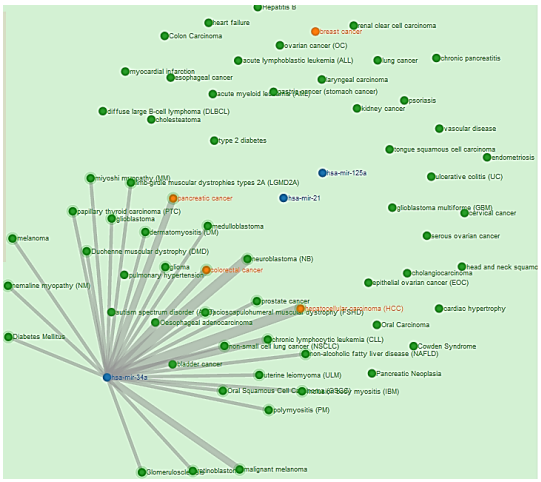


Fig. 18.: miRNA's range of influence

Whereas, breast cancer (see black dotted arrow in Figure 17) which was one of the resultant diseases in the MWM approach, is listed further below the aforementioned diseases with 265 motifs (observe its placement in the Figure 17) since its being regulated by two miRNAs and thereby less motifs. Upon clicking a certain miRNA, its range of influence can be observed as displayed in Figure 18.

Using the visualization, the user can also determine paths or shortest paths to unrelated diseases, for e.g. see Figure 19, the disease cholagiocarcinoma and melanoma seem to be unrelated. However, upon drawing careful egdes, it can be noted that melanoma is three hops away from cholangiocarcinoma, via papillary throid carcinoma (PTC). Upon the activation of the disease PTC, mir-34a and mir-21 are active and thereby the weak possibility of the activation and association of cholangiocarcinoma and melanoma. Similar such paths between diseases of interest can be explored by the user.

It is important to note that, this visualization does not provide strong certainty in predicting or determining the disease-disease interaction, rather merely provides the abstract idea of the reach of the diseases onto each other. However, this tool *does*

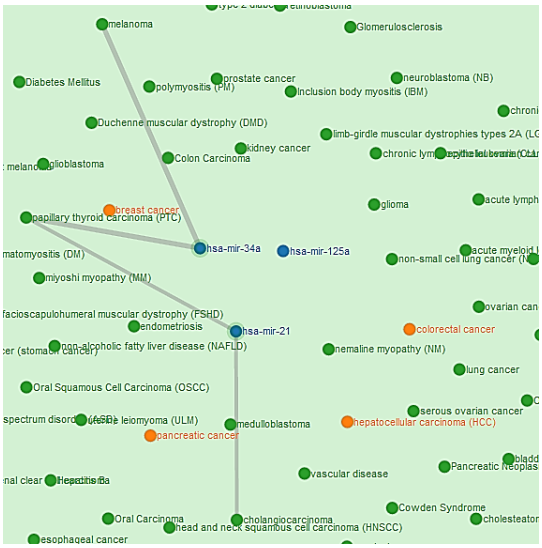


Fig. 19.: Paths between diseases

provide the preliminary overview of the disease-disease interaction network which can be studied adeptly to uncover significant underlying associations.

The above methodology was published in Nalluri et al.,2015[39]

Table 1.: MWM based algorithm results

S.No.	**miRNAs**	**Diseases**	**PubMeds for results**
1	hsa-mir-9-1, hsa-mir-9-2, hsa-mir-200c	Breast cancer, Colorectal cancer, Kidney cancer, Ovarian cancer	23617747
2	hsa-mir-182, hsa-mir-200a, hsa-mir-200b, hsa-mir-200c	Lung cancer, Ovarian cancer (OC), Hepatocellular carcinoma (HCC), Breast cancer, Kidney cancer, Colorectal cancer, Oral squamous cell carcinoma	23272653
3	hsa-mir-29a, hsa-mir-34a, hsa-mir-34b, hsa-mir-25	Ulcerative coltis, Serious ovarian cancer, Bladder cancer, Pituary adenoma, Primary Biliary cirrhosis, Epithelial Ovarian Cancer, Cardiac hypertrophy, Breast cancer, Acute Lymphoblastic leukemia, Kidney cancer, Gastric cancer and nasopharyngeal carcinoma	18056805, 19475496, 19646430, 16461460, 18728182, 18390668, 17823410

Table 2.: Example for disease participation in square motifs with the set of input miRNAs using motif analysis

Input miRNAs	**Disease (participation count in motifs)**
hsa-mir-184 hsa-mir-200a hsa-mir-200b hsa-mir-200c	Malignant melanoma (37) Epithelial ovarian cancer (EOC) (37) Breast cancer (37) Lung cancer (37) Cancer (23) Ovarian cancer (OC) (23) Serous ovarian cancer (23) Hepatocellular carcinoma (HCC) (18) Kidney cancer (18) Endometriosis (9) Non-alcoholic fatty liver disease (9) Oral squamous cell carcinoma (OSCC) (8) Colorectal cancer (5)

CHAPTER 5

CONSENSUS BASED NETWORK INFERENCE

The multi-level interactions of miRNAs in cancer-like multi-factor diseases are more complex due to the possibility of several types of interactions such as the classical miRNA-mRNA, miRNA-environmental factors, miRNA- transcription factors-miRNA, and our newly hypothesized direct miRNA-miRNA interactions without any intermediate linkers such as transcription factors. However till date, no experimental proof of direct miRNA-miRNA interactions exists except a single study reported in mouse[65].

Although, the patterns and reasons behind miRNA's deregulation in cancers are not fully understood, it has been found that miRNAs operate in clusters and tend to work together in groups,[66]as evidenced in certain diseases[20]. Such coordinated regulation, mutual co-targeting and co-regulation, and miRNA regulation by other miRNAs are reported in many disease conditions including various cancers[66]. To elucidate the miRNA-disease associations at regulome level, we developed the *miRegulome* database and analysis tools[51]. Furthermore, in cancers it has been observed that groups of miRNAs, known as *'superfamily'*, express consistently across several cancers and may act as *drivers* of tumorigenesis, where few key miRNAs direct the global miRNA expression patterns[67]. Identification of such groups of super-families of miRNAs obviously leads to the intuition that the therapeutic suppression or expression of any one the miRNAs in the family would compensate for the other participants of the family[67]. Our hypothesis is that, these miRNAs in such 'super-families' may interact directly or indirectly, by forming a core miRNA-miRNA co-regulatory net-

work and thereby acting as a signature component for prognosis, prediction, and early diagnosis of any disease including cancer.

5.1 Methods

In this work, we have used the miRNA expression data sets available in the *PhenomiR*[68] database to predict miRNA-miRNA core/signature interactions across several cancers using a combination of (i) six state-of-the-art network inference algorithms, (ii) a *wisdom of crowds*[69] based consensus approach[70] to generate disease-specific miRNA interaction networks with higher accuracy, and (iii) a simplified graph intersection analysis to identify the miRNA-miRNA core interactions across multiple diseases belonging to a particular disease class.

The methodology adopted in this paper is comprised of i) translating the miRNA-disease expression scores from the *PhenomiR* database into a miRNA expression matrix (Figure 20, Step 1); ii) deploying six network inference algorithms on the expression matrix and deriving the miRNA-miRNA interaction scores from each algorithm (Figure 20, Step 2); iii) performing a consensus-based approach, i.e. estimating an average score for every miRNA-miRNA interaction across its six predicted scores (Figure 20, Step 3); iv) validating the resultant interactions using precision-recall analysis with a hypothetical true network generated using the PubMed IDs from *PhenomiR*; v) analyzing the miRNA-miRNA interaction networks for every disease and detection of the conserved miRNA-miRNA interactions across various groups of cancers and finally vi) validating the conserved miRNA-miRNA interactions in the identified group of cancers via manual literature search.

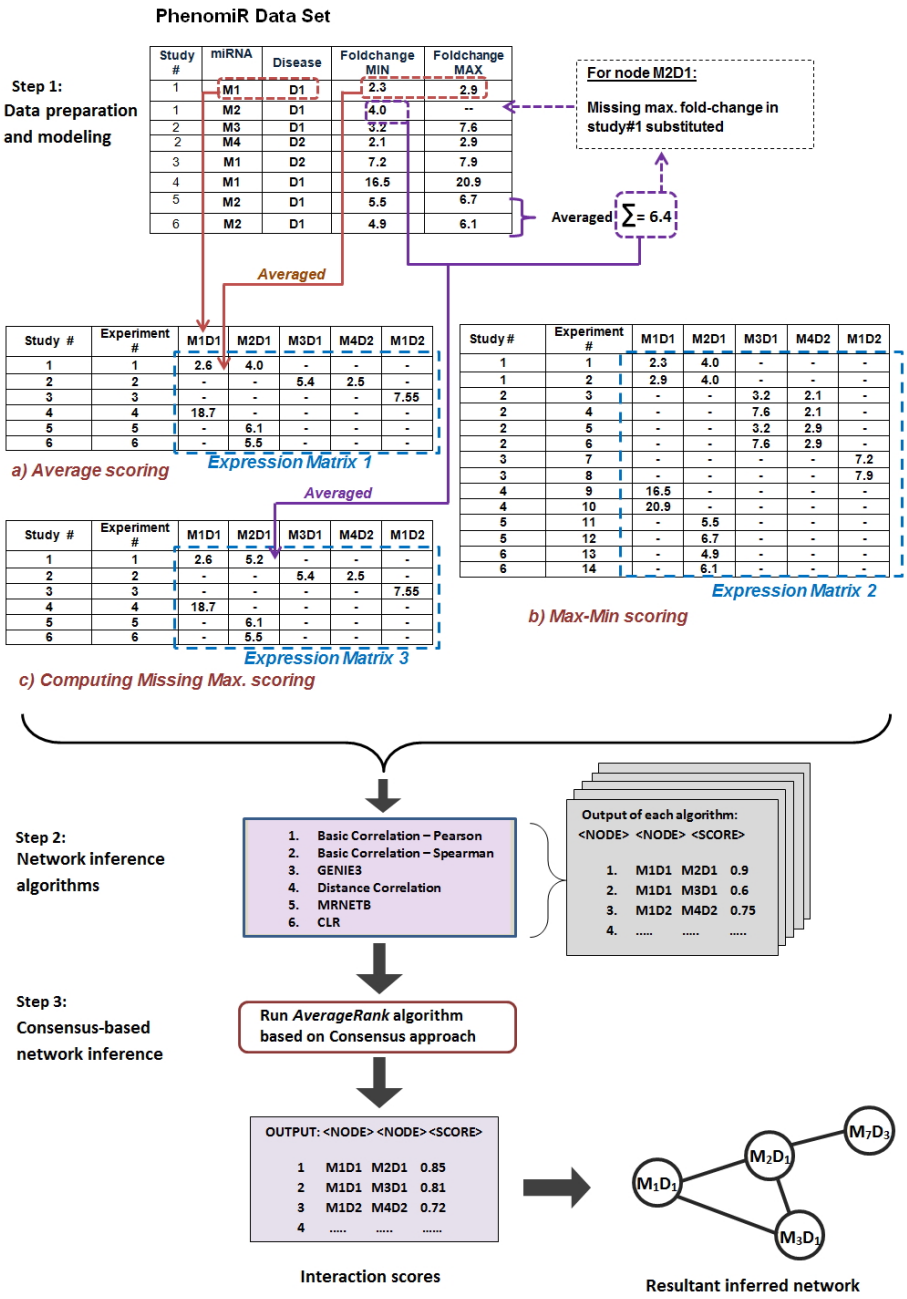


Fig. 20.: Overview of the methodology with M_i denoting miRNAs and D_j denoting the diseases. *Step 1* consists of translating the *PhenomiR* data set into three miRNA expression matrices (a, b and c) based on three approaches. In *Step 2*, each of these matrices are subjected to six network inference algorithms which produce the interaction scores across the different M_iD_j nodes. In *Step 3*, the six individual $M_iD_j - M_xD_y$ interaction scores are averaged into a final score designating its confidence.

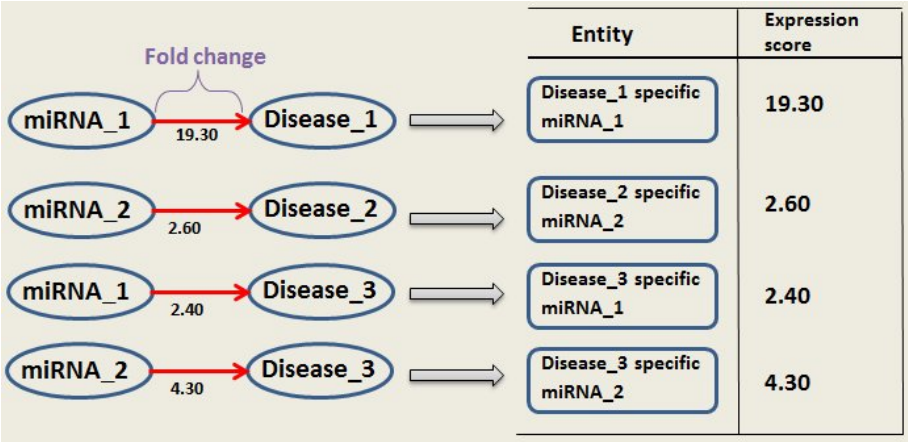


Fig. 21.: Schematic of the miRNA-disease regulation with fold-change values.

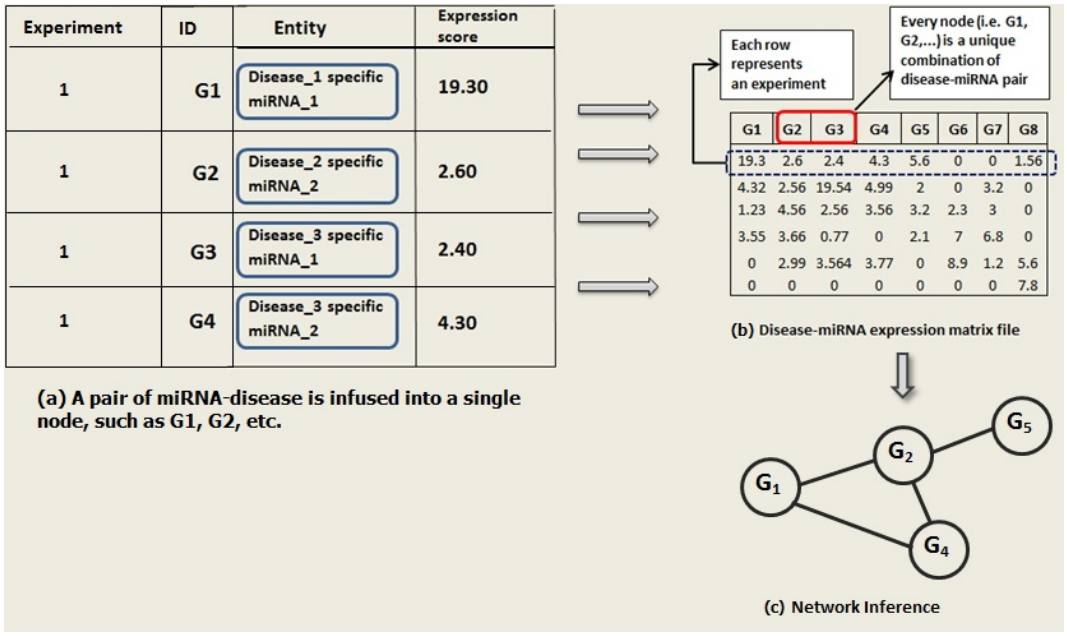


Fig. 22.: Schematic of the miRNA expression data set. [(a) and (b)]: Data from *PhenomiR* is mapped into an miRNA expression matrix. (c) Network inference approach is applied to the matrix to derive the probabilistic interaction network.

5.1.1 Data preparation and modeling

The data from the *PhenomiR* database is freely available and was used in this study. *PhenomiR 2.0* was downloaded for the purposes of this study. *PhenomiR 2.0* is a comprehensive data set containing 535 database entries across 345 articles recording miRNA expressions in diseases[68]. As shown in Figure 21, the data from *PhenomiR* was converted into a disease-specific miRNA expression matrix (shown in Figure 22). The miRNAs whose fold-change values were not available in *PhenomiR 2.0* data set were discarded from the study; this also includes some misformatted lines of data that were excluded from further processing as they were also missing the fold-change values. Here, the core idea is to consider a pair of miRNA and disease as a single miRNA-disease (*MD*) node, as seen in Figure 22; note that, for ease of reference, we consider an M_iD_j pair as an *MD* node which conceptually designates a disease-specific miRNA. The same miRNA participating in multiple diseases will have different expression profiles in each of them and hence the disease specific miRNA terminology, i.e., *MD*, signifies a miRNA's expression profile in a particular disease. Thus, every unique miRNA-disease pair constitutes a unique *MD* type node. In this disease-specific miRNA expression matrix (Figure 22-b), each row represents a study/experiment and each column represents an *MD*'s expression score in that study. The resultant expression matrix herein, has 4,343 unique nodes/columns (i.e., unique *MD*s in the network) for 267 samples (i.e., rows).

In the *PhenomiR* data set, some *MD*s have two fold-change values indicating minimum and maximum expression scores while other *MD*s only report a minimum fold-change expression score (for e.g., see Figure 20, Step 1, *PhenomiR* data set, row 2). To assess these scenarios, we devised three different methodologies (described in the next section), generated separate expression matrices based on each methodology

and performed the subsequent analysis on each of them.

Average scoring

Under *Average scoring* method, for the *MD*s having both minimum and maximum fold-change values per sample, their average was taken and considered as the final expression value in the expression matrix. As shown in Figure 20, Step-1, the entry $M1$-$D1$ has two expression values - 2.3 and 2.9, i.e., minimum fold-change and maximum fold-change respectively, which were averaged to 2.6 in the *Expression Matrix 1* (see Figure 20, Step 1-a). For *MD*s with only their minimum expression values reported, this single value was also considered to be its average expression value.

Retaining maximum and minimum expression values

The *Average scoring* method can lead to a potential loss of information as the individual maximum and minimum expression values (when available) were not retained. Hence we designed the following two methods to generate the expression matrix.

1. *Max-Min scoring*

   Under *Max-Min scoring* method, for the *MD*s having minimum and maximum fold-change expression values, (instead of taking their average) both these data points were considered as separate entries; thus, the same *MD* was considered twice in the expression matrix with the duplicate entry designating a new experiment. As displayed in Figure 20, Step 1, the first row entry, $M1$-$D1$ in *Study-1* has two expression values; these values were individually considered as separate data points and included in the expression matrix accordingly along with their co-expressing miRNAs' expression values, providing us with *Expression Matrix 2* (see Figure 20, Step 1-b).

2. *Computing Missing Max. scoring*

   Under *Computing Missing Max scoring* method, for the *MD*s which did not have a maximum fold-change expression value, we took an average of its maximum fold-change values across all its *other samples* and substituted this average score as it's maximum fold-change expression value. As shown in Figure 20, Step 1, the entry $M2$-$D1$ on 2nd row does not have a maximum fold-change value. However, $M2$-$D1$ combination has maximum fold-change expression values of 6.7 and 6.1 from sample #5 and #6, respectively. Herein, we took an average of these two values, i.e. 6.4 and substituted it for the original missing value for $M2$-$D1$ in the 2nd row. This method overcomes the limitation posed due the non-availability of the expression value by giving its closest approximation, based on the particular *MD*'s expression pattern across the sample spectrum. After applying this method, the *Average Scoring* method was performed on this matrix to obtain *Expression Matrix 3* (see Figure 20, Step 1-c).

After the three expression matrices were derived, a reverse engineering methodology[70] was adopted to reconstruct the *MD-MD* regulatory network from these expression matrices (Figure 22, *Network Inference*), by applying six widely used network inference algorithms along with a consensus-based ranking algorithm, which is explained in the next section.

5.1.2 Network inference algorithms

Each expression matrix has 4,343 nodes and therefore, there are potentially 4,343 x 4,343 (i.e. 18,861,649) *MD-MD* interactions in the network. Six different network inference algorithms were applied on the miRNA expression matrix, which gave prediction scores for every *MD-MD* interaction. We used the mutual information-based algorithm, Context Likelihood of Relatedness (CLR)[71], Maximum Relevance Min-

imum Redundancy Backward (MRNETB)[72], Basic Correlation methods (Pearson and Spearman), Distance Correlation (DC)[73], and regression-based Gene Network Inference with Ensemble of Trees (GENIE3)[74] algorithms for network inference. The details of the algorithms are given in *Appendix A*. Note that, the Basic Correlation methods resulted in two different network inference algorithms based on the type of correlations implemented, i.e., one each for Pearson and Spearman correlations.

5.1.3 Consensus based network inference approach

Each of the six individual network inference algorithms produced a ranked list of prediction scores for every *MD-MD* interaction (see Figure 20, Step-2). Thereafter, we used the *wisdom of crowds*[69] approach, which proposes that the aggregation of information from the community yields better results than the individual few. In this study, the consensus based approach aggregates the collective information (i.e. prediction scores) from the six individual network inference algorithms and computes a more accurate final score for *MD-MD* interactions. This rank is computed by taking an average of the predicted ranks of each interaction derived from the corresponding network inference algorithms. Figure 23 displays the workflow of this approach. This approach was earlier implemented to infer gene-regulatory networks and yielded highest accuracy compared to each of the individual network inference algorithms[70].

This consensus based network inference approach is executed in the *Average Rank*[70] algorithm which essentially computes the average score of a particular *MD-MD* interaction by taking the mean of its six predicted ranks. The ranking methodology used in this algorithm is based on the *Borda* count method. This method is used in elections during which voters rank candidates as per their preferences. The winning candidate is the one with the best average rank. Here, all the interactions

are first ranked in descending order of their predicted scores (as seen in the column *Rank* in Table 3). Describing briefly, the *Borda* count method allocates points to each rank. The highest ranked interaction (meaning, 1) get the maximum *Borda* points (number of interactions - 1) and the lowest ranked interaction has 0 *Borda* points as demonstrated in the column *Borda points* in Table 3. In order to derive the final rank between 0 and 1, these points are thereafter normalized to derive a relative *Borda* rank. Thus, each rank has been translated to its new relative *Borda* rank. Note that, the *Borda* count ranking method is among the many other methods to perform *averaging* of the ranks in the consensus methodology.

The six network inference algorithms generate six different ranks for each interaction and the consensus algorithm next computes an average *Borda* rank for the interaction. Tables 3 and 4 display a scenario of ranking four *MD-MD* interactions I_1, I_2, I_3 and I_4 via a consensus-based approach as executed in *AverageRank* algorithm. Table 3 displays the ranked list of predictions for these interactions by all the six network inference algorithms based on their prediction scores. For example, in Table 3, *Algorithm 1* ranks *MD-MD* interactions in this order — I_4, I_2, I_1 and I_3 based on their prediction scores. The individual ranks for miRNA-miRNA interaction I_4 are 1, 3, 4, 2, 2 and 3 by the six algorithms respectively (as highlighted in blue), and their relative respective *Borda* ranks are 1, 0.333, 0, 0.666. 0.666 and 0.333. The final rank of interaction I_4 is the average of all the *Borda* ranks, i.e., 0.49, as demonstrated in Table 4 (as highlighted in blue). Similarly the final ranks of every other interaction is computed using the following formula,

$$Final - rank(I) = \frac{1}{K} \sum_{j=1}^{K} Borda - rank_j(I) \tag{5.1}$$

where, K is the number of algorithms (six, in our case). These results are

displayed in Table 4.

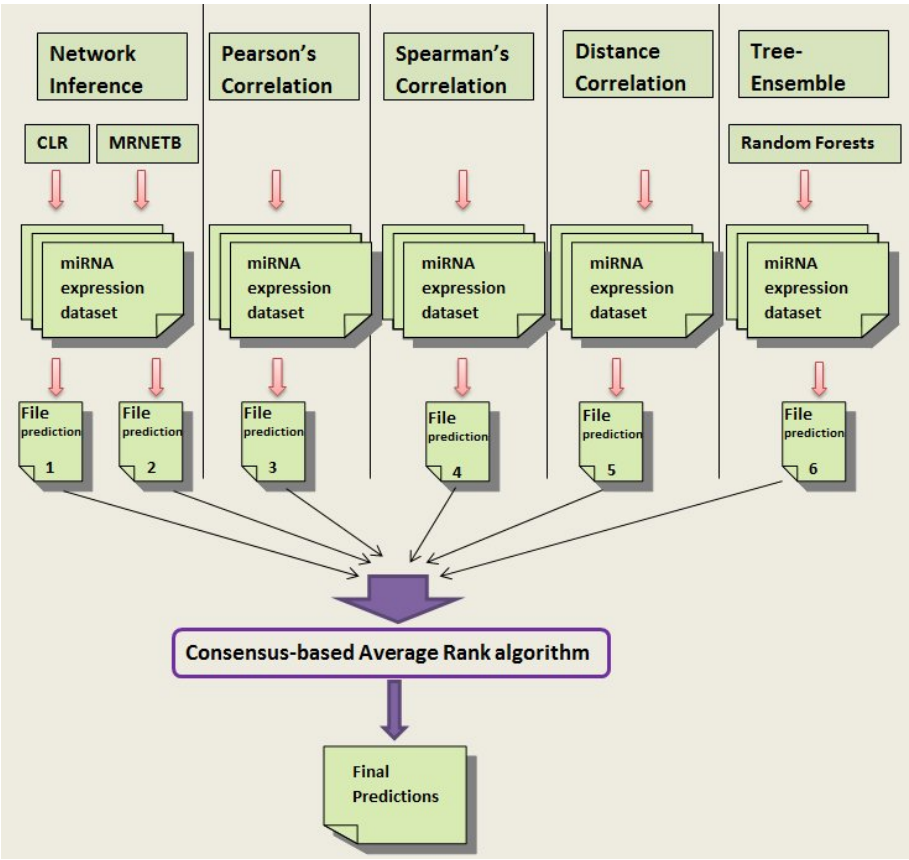


Fig. 23.: Workflow of the consensus-based miRNA network inference.

An example of the final result listing of our *MD-MD* interactions is shown in Table 5 (also see Figure 20 Step-3):

In these results, we noted all the different possibilities of interactions that can occur considering the miRNA-disease pair, i.e. *MD* as a node. There are essentially four types of interactions that can exist in this network. These are explained in Table 6. Among these types, *type 1* is a self-loop and not applicable for our purposes. For application purposes of our methodology, we focused on analyzing the set of interactions belonging to *type 3* which is further elaborated in the next section. Interactions

of *type 2* and *type 4* will be studied in the future to analyze the relationship between diseases sharing a common miRNA (*type 2*) and the proximity between dissimilar miRNAs and dissimilar diseases (*type 4*) having high probabilities of interaction.

Disease-specific miRNA network construction

In this section, the results of the *type 3* interactions were selected for disease-specific analysis. There were 66 unique diseases in the final predicted list of interactions from the *Average Rank* algorithm. Under a specific disease D_x, all the miRNA-miRNA edges, i.e. M_1D_x-M_2D_x edges were collected into a single D_x disease network; thereby giving us the **d**isease-specific **m**iRNA-miRNA **i**nteraction **n**etwork (*DMIN*) (Figure 24, Step 1). *DMIN* is a network *G = (V, E)*, where $V = \{M_1D_x, M_2D_x, ...M_nD_x\}$ (i.e., set of miRNAs under disease name D_x) and E is the ordered set of edges, where edge $e = \{M_i, M_j\}$. We performed a similar network construction for every cancer-related disease, D_x. To pursue a more definitive and cancer-specific analysis, only cancer-related diseases were chosen and grouped into classes based on their tissue/organ specificity. We created four major classes: i) gastrointestinal cancers (esophageal, gastroesophageal, gastrointestinal, gastric, and colorectal cancer), ii) endocrine cancers (hepatocellular, pancreatic, and thyroid carcinoma follicular, and thyroid carcinoma papillary), iii) leukemia/blood cancers (hematological tumors, acute myeloid leukemia, chronic lymphatic leukemia, and acute myelogenous leukemia), and iv) nerve cancers (neuroblastoma, medulloblastoma, and glioblastoma).

Under a particular disease class, all the corresponding *DMIN*s were combined into a single network (Figure 24, Step 2). Using graph intersection analysis, we mined the miRNA-miRNA interaction networks of all the cancers within the specific class to identify a conserved (signature/core) miRNA-miRNA interaction component. This identified miRNA-miRNA interaction component was present in all the diseases of

that particular class. These findings are reported in the Section *pan-cancer miRNA signatures* and the results are discussed in the Section *Discussion*.

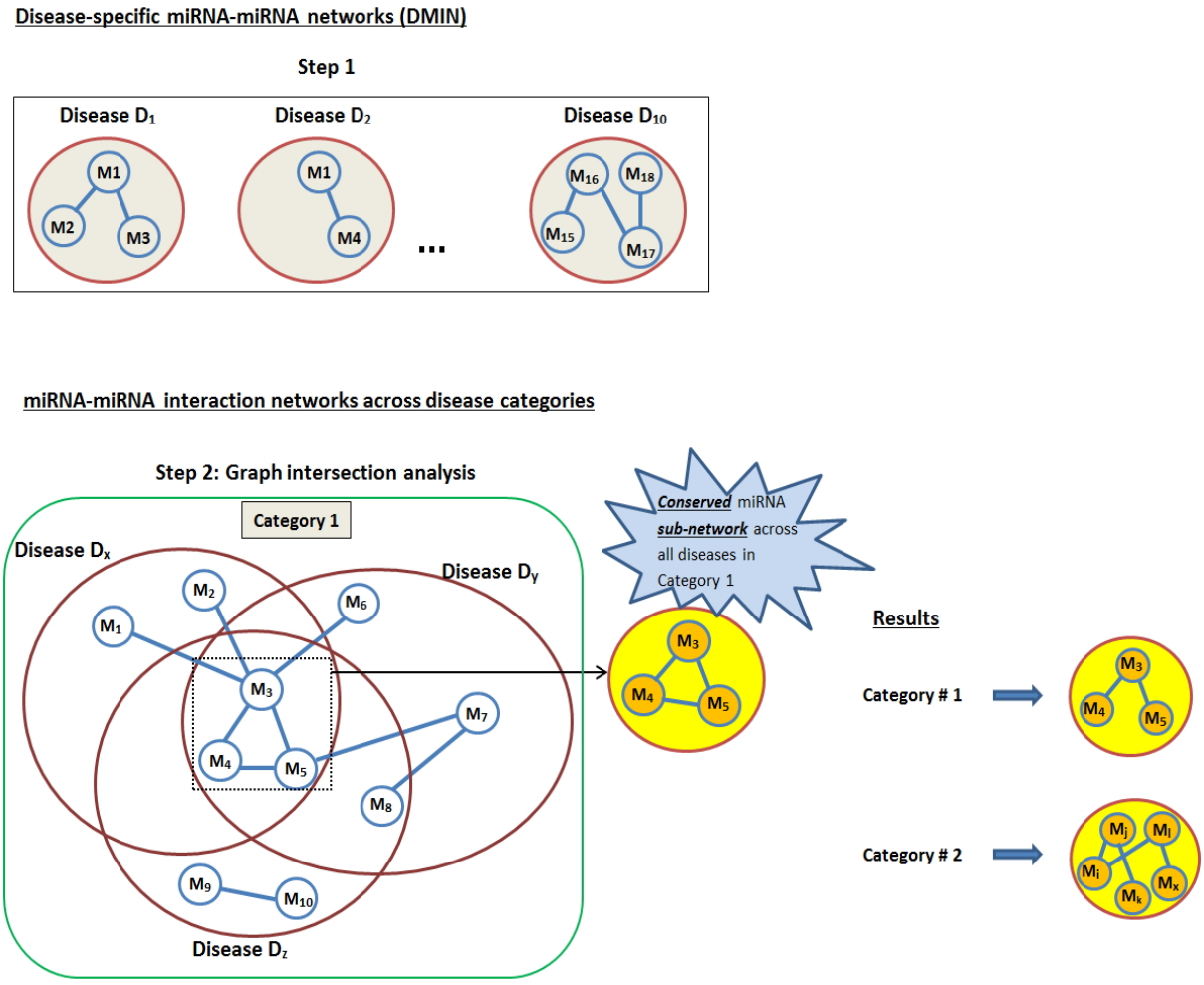


Fig. 24.: Overview of the disease analysis.

5.2 Results

Validation of interactions

After executing the *Consensus based network inference approach* on three input miRNA expression matrices derived from the three approaches mentioned in the *Data preparation and modeling* section (*Average scoring*, *Retaining Max-Min* and *Comput-*

ing Missing Max.), we obtained three sets of predicted miRNA-miRNA interactions. Each predicted interaction was validated by querying for PubMed IDs in the *PhenomiR* database which cited and reported the occurrence of miRNAs' association with the specific disease in a single PubMed ID. For e.g., for each predicted interaction, i.e. M_aD_x to M_bD_x, if a PubMed ID cited the occurrence of the association between the miRNAs (M_a, M_b) and the disease (D_x), the interaction was termed as true/-validated (1); else the predicted interaction was termed as unknown/unverified (0). Based on this, labels were generated for every interaction in the resultant set forming the true network. We performed a precision-recall analysis to ascertain the accuracy of the consensus-based network inference method. The precision-recall values were calculated using the formula:

$$Precision = \frac{tp}{tp + fp} Recall = \frac{tp}{tp + fn} \tag{5.2}$$

where tp, fp, and fn are true-positives, false-positives and false-negatives respectively.

Figure 25 displays the results of the precision-recall analysis and the ROC curve for all the three approaches used. As demonstrated in the figure, the *Average scoring* method fared better than the other two methods; in fact the *Computing Missing Max.* method also performed well for low recall but gradually degraded for higher recall values. Based on this precision-recall curve, our proposed methodology displays a high precision (for up to a 30% recall) demonstrating its effectiveness in providing high confidence to the results. The ROC curve shows that both the *Average scoring* and *Computing Missing Max.* methods are comparable in predicting the true positives when compared to the number of false positives seen alongside.

Note that our true network generation method has some obvious limitations. While a true edge constituting the association of the two miRNAs with the same

disease in the same PubMed ID is still acceptable (specifically because these edges were manually curated), the unverified edges may simply mean that a study has not yet been reported associating the miRNAs to the same disease. Hence, a high precision performance should be the best judge of our methodology whereas the recall curve can be somewhat circumstantial.

Pan-cancer miRNA signatures

After the *Validation of interactions*, in order to confidently detect miRNA signatures in the specified disease classes, only the top 10% interactions with the highest confidence scores were used in the construction of *DMIN* (Figure 24, Step 1) and the subsequent graph intersection approach (Figure 24, Step 2). Hence, all the considered miRNA-miRNA interactions had a confidence score of 0.9 and above. As reported in Figure 26, under gastrointestinal cancers, we detected a signature component of three miRNAs (hsa-mir-30a, hsa-mir-181a-1, and hsa-mir-29c). For endocrine cancers, the signature component consisted of hsa-mir-221, hsa-mir-222, hsa-mir-155, hsa-mir-224, hsa-mir-181a-1, and hsa-mir-181b-1. For leukemia cancers, the signature component consisted of hsa-mir-29b-1, hsa-mir-106a, hsa-mir-20a, hsa-mir-126, and hsa-mir-130a. We observed two different signatures for nerve cancers. For subsequent validation of these cancer-specific signature set of miRNAs, we manually mined PubMed articles which corroborate our results, as reported in Figure 26. We queried both the *PhenomiR* database and the *PubMed Central* database for these reported PubMed IDs; the results from these two sources are shown in different colors in Figure 26. We also observed that, while hsa-mir-30 is common in gastrointestinal and nerve cancers; hsa-mir-181 is shared by gastrointestinal, endocrine and nerve cancers. The miRNA signature component of the category leukemia is found to possess a distinct group of miRNAs (Figure 26). The role and involvement of these miRNAs in their

associated diseases are further elaborated in the *Discussion* section.

The individual steps involved in the manual search process from *PubMed Central* are shown in Figure 27. To summarize, we first searched *PubMed Central* with the list of core miRNAs and each disease for which they form a signature component. We next manually checked the 'search' results to confirm the associations (i.e., the pruning step for PMIDs). If not enough results were retrieved from this search, we entered each miRNA, disease pair individually for all the miRNAs forming the signature component in that disease; each of these results were then manually pruned and collated to give us the set of PMIDs corresponding to the core miRNAs for that disease. This process was repeated for all the other diseases of a particular disease class.

5.2.1 *miRsig* - an online tool

In order to aid researchers to identify disease-specific miRNA-miRNA interaction networks across several diseases, we developed the *miRsig* tool, available at *http://bnet.egr.vcu.edu/miRsig*. *miRsig* allows the user to visualize the miRNA-miRNA interaction network for each disease recorded in *PhenomiR* and also across multiple diseases. The results are based on the consensus-based network inference approach. *miRsig* also allows users to search for a common/core miRNA-miRNA interaction component in a user-specified selection of diseases (see Figure 28). Users can create their own class/category of cancers by *selecting* more diseases, as shown in Figure 28. The edges in the interaction have confidence scores as weights, from 0 (minimum) to 1 (maximum). Hence, the tool also allows the user to view only the higher/lower/specific confidence interactions by changing the *Maximum* and *Minimum* confidence score ranges. Currently, the total number of edges across the entire miRNA-miRNA interaction networks are more than 18 million. Hence, to avoid clut-

tering of the result set and to allow clear visibility and comprehension of the network, the *Minimum* score is set to 0.5, if not specified by the user. Users can also view and analyze the topological properties of miRNA clusters interacting in each or a set of diseases. The signature/core miRNA-miRNA interactions among esophageal, gastroesophageal, gastrointestinal, gastric, and colorectal cancers, as predicted and visualized is shown in Figure 28. This network component consisting of three miRNAs (has-mir-30a, has-mir-181a-1, and has-mir-29c) is the signature component for all the aforementioned five cancers, and can be validated using simple literature search on *PubMed Central* database as demonstrated in Figure 27.Users can also download the miRNA-interaction network in the format of an edge-list in a CSV file. This edge-list can be imported in various network analysis tools such as, *NodeXL, Cytoscape*, etc. for further study and analysis of the interaction network.

miRsig tool has been developed using MySQL as the back-end database and HTML, PHP, JavaScript, AJAX for front-end design. The interactive network visualization has been implemented using data visualization library, D3.js[75].

5.2.2 Discussion

miRNA-mRNA interactions have been substantially documented [76] and is a prime area of ongoing research. Similarly, miRNA- miRNA interactions through mutual co-expression[77], via transcription factor[78], and miRNA-disease associations[79] have also been reported. However, miRNA-miRNA interactions towards identification of a core miRNA-miRNA module that could potentially be a signature component for a particular disease have not been studied enough. Many studies have used computational approaches to study this aspect. A miRNA-miRNA co-regulation network in lung cancer was identified using a progressive data refining approach[35]. Similarly, miRNA expression profiling along with a genome-wide SNP approach was

used to create a miRNA-miRNA synergistic network to study coronary artery disease[80]. miRNA-miRNA interactions were also identified in esophageal cancer using K-clique analysis on a bipartite network consisting of miRNAs and subpathways[81]. Additionally, miRNA-target interactions were integrated with miRNA and mRNA expressions to deduce miRNA-miRNA interactions in prostate cancer[82]. A network topological approach was also undertaken to identify disease miRNAs by constructing a miRNA-miRNA synergistic network consisting of co-regulating functional modules[34].

In this work, we adopted a strategy that takes a miRNA expression profile and uses six different network inference algorithms (CLR[71], MRNETB[72], Basic Correlation (Pearson and Spearman), DC[73], GENIE3[74]), each varying in their inference strategies, integrated with a consensus approach and graph intersection to identify the conserved miRNA-miRNA interaction signature across a group of diseases (cancers, in this case). The identified signatures were validated via manual literature search and were found to be associated within the classes of the selected cancers, demonstrating the efficacy of the method. Under validation, we retrieved the PMIDs reporting the associations from the *PhenomiR* database and also performed a manual literature search in the *PubMed Central* database to separately corroborate our results, as displayed in Figure 26.

Our results show that, the expression profile of hsa-mir-30a, hsa-mir-181a-1, and hsa-mir-29c could be a signature for gastrointestinal cancers that comprises of esophageal, gastroesophageal, gastrointestinal, gastric, and colorectal cancers (Figure 26). These miRNAs are already reported to be associated with these cancers[83–86]. miRNAs (hsa-mir-30a, hsa-mir-29c, hsa-mir-181a-1) displayed the same trend of expression in a study of esophageal adenocarcinoma (EAC) and Barrett's esophagus (BE) and were differentially up-regulated in both the disease tissues. hsa-mir-181a

and hsa-mir-29c showed higher expression levels in EAC to that of BE with high grade dysplasia[86]. Studies have also reported hsa-mir-181a, hsa-mir-30a and hsa-mir-29c being overexpressed in esophagela carcinoma (EC) and hsa-mir-29c to be underexpressed in EC[87][88] and therefore, this group of miRNAs may be considered for developing a pan-diagnostic tool for the aforementioned cancers.

We identified that hsa-mir-221, hsa-mir-222, hsa-mir-155, hsa-mir-224, hsa-mir-181a-1, and hsa-mir-181b-1 make the signature for endocrine cancers (hepatocellular, pancreatic, and thyroid cancers) (Figure 26). Reports suggest that these miRNAs are predominantly associated with this group of cancers[89–92]. In another study analyzing molecular signatures for aggressive pancreatic cancer, all the miRNAs (hsa-mir-221, hsa-mir-222, hsa-mir-155, hsa-mir-224, hsa-mir-181a-1, and hsa-mir-181b-1) were significantly altered due to chronic exposure to conventional anti-cancer drugs[93]. A large-scale meta-analysis investigating candidate miRNA biomarkers for pancreatic ductal adenocarcinoma (PDAC) across eleven miRNA expression profiling studies, reported all the miRNAs to be up-regulated and having a consistent direction of change. miRNAs hsa-mir-221, hsa-mir-222, hsa-mir-155 were reported to be upregulated together in at least five of these studies with a consistent direction. Among them, miRNAs hsa-mir-221, hsa-mir-155 were identified as part of a meta-signature and biomarkers for PDAC[94]. Studies also report all these miRNAs to be associated with lung cancer[95]. Thus this set of miRNAs may be used/tested as a diagnostic tool for all the endocrine cancers considered here.

Seven miRNAs (hsa-mir-29b-1, hsa-mir-146a, hsa-mir-20a, hsa-mir-126, hsa-mir-99a, hsa-mir-199b and hsa-mir-130a) that are well documented for their association with various kinds of leukemia[92, 96–101] are found to form the signature component of leukemia from our analysis (Figure 26). miRNAs (hsa-mir-29b-1, hsa-mir-20a, hsa-mir-126, hsa-mir-146a, hsa-mir-199b) were differentially expressed in a blood stem

cell study in which the blood stem cells were treated with plerixafor and granulocyte colony-stimulating factor. The miRNAs were recorded to be expressed in this treated cell study analyzing acute lymphocytic leukemia conditions[102]. miRNAs (hsa-mir-126, hsa-mir-130a, hsa-mir-99a, hsa-mir-146a, hsa-mir-199b) have also been reported to express together in a myeloid cell study exploring transcription factor binding site motifs[103]. Therefore, this signature group of miRNAs can be potentially used as a screening or diagnostic tool for a range of different types of leukemia.

In case of neurone cancers (neuroblastoma, medulloblastoma, and glioblastoma) we detected two signatures: i) hsa-mir-323, hsa-mir-129-1, hsa-mir-137, hsa-mir-330, hsa-mir-149, hsa-mir-107, hsa-mir-30c-1, hsa-mir-181b-1 and ii) hsa-mir-30b, hsa-mir-331, hsa-mir-150, hsa-let-7a-1 (Figure 26). Regarding the first signature network component, hsa-mir-137, hsa-mir-330, hsa-mir-149, hsa-mir-107, hsa-mir-181b were among the miRNAs whose experimentally validated targets (such as CTBP1, CDC42, CDK6, E2F1, VEGFA, AKT1, KAT2B) affect the pathways which play a crucial role in glioblastoma biology. Deregulations of hsa-mir-137, hsa-mir-330 and hsa-mir-149 lead to effects in the glioma de novo pathway, VEGF signaling pathway and Notch signaling pathway [104]. Among the miRNAs reported in the second signature component, hsa-mir-330 and hsa-mir-30b are among the top ten miRNAs having least coefficient of variation in the expression of benign kidney tumor and hsa-mir-150 is differentially expressed in metastatic clear cell renal cell carcinoma[105].

Comparing our results with other similar works has been challenging, primarily because there are not many studies that have reported direct miRNA-miRNA co-regulations across these disease classes. Similar studies[35, 67, 106] have used different disease and miRNA data sets which makes a one-to-one comparison challenging. In some previous works, miRNA-miRNA regulatory associations have been deduced based on the semantic similarities between the associated diseases[25] and

based on the analysis of shared transcription factors, common targets, KEGG pathway analysis and corroboration from literature[35]. However, none of these methods allow for a network-level miRNA-miRNA analysis for a variety of diseases and hence cannot be used for comparison purposes to the predicted interaction networks in this paper.

Online analysis and visualization of results is an aid to the research community. Along these lines, several network analysis and visualization tools have been developed, such as *VisANT* for integrative online visual analysis of biological networks and pathways[107], *miRegulome* for miRNA regulome visualization and analysis[51] and *miRNet* for functional analysis of miRNAs within a high-performance network visual analytics system[108] among others. However, no tool is available so far which can perform an online visualization and analysis of signature miRNAs across multiple diseases. The *miRsig* tool developed here bridges this gap and provides an intuitive analysis and visualization of core/signature miRNA-miRNA interaction components for several diseases.

In this work, we have developed a powerful consensus-based network analyses to identify disease specific miRNA-miRNA interactions. The method is effective in identifying the signature/core miRNA-miRNA interactions for a group of diseases; here tested on cancer. These signature miRNAs would have potential use as diagnostic/ prognostic/therapeutic values in a group of related diseases such as cancers. *miRsig* is a powerful prediction and visualization tool for core/signature miRNA-miRNA interaction among a number of user specific diseases.

This work was published in Nalluri et al.,2017[109]

Borda rank **(Norm. borda points)**	**Borda points**	**Rank**	**Alg. 1**	**Alg. 2**	**Alg. 3**	**Alg. 4**	**Alg. 5**	**Alg. 6**
1	3	1	I_4	I_2	I_2	I_2	I_2	I_3
0.6667	2	2	I_2	I_3	I_3	I_4	I_4	I_2
0.3334	1	3	I_1	I_4	I_1	I_3	I_3	I_4
0	0	4	I_3	I_1	I_4	I_1	I_1	I_1

Table 3.: Ranked individual predictions of each algorithm for every interaction I. **Borda points** are allocated to each **Rank**. A relative ***Borda rank*** ($= \frac{Borda\ points\ for\ that\ rank}{maximum\ Borda\ points}$) is computed for every **Rank**. *Borda* ranks for interaction I_4 (highlighted in blue) are 1, 0.333, 0, 0.666, 0.666 and 0.334 by the six algorithms respectively.

Interaction	**Averaging of *Borda* ranks**	**Final rank**
I_2	(0.66 + 1 + 1 + 1 + 1 + 0.66) / 6	0.88
I_3	(0 + 0.66 + 0.66 + 0.33 + 0.33 + 1) / 6	0.49
I_4	(1 + 0.33 + 0 + 0.66 + 0.66 + 0.33) / 6	0.49
I_1	(0.33 + 0 + 0.33 + 0 + 0 + 0) / 6	0.11

Table 4.: Final ranks for each interaction; the final rank of interaction I_4 is 0.49

Rank	**Interaction**	**Score**
1	Hepatocellular carcinoma:hsa-mir-183 ⇒ Hepatocellular carcinoma:hsa-mir-374a	0.9786
2	Hepatocellular carcinoma:hsa-mir-374a ⇒ Hepatocellular carcinoma:hsa-mir-182	0.9781
3	Breast cancer:hsa-let-7a-1 ⇒ Breast cancer:hsa-mir-30d	0.2985
4	Breast cancer:hsa-let-7a-1 ⇒ Breast cancer:hsa-mir-381	0.2426

Table 5.: Format of the results based on the consensus approach.

Type #	**Interaction type**	**Edge**	**Remark**
1	$miRNAs_{same}, Diseases_{same}$	$M_1D_1 \rightarrow M_1D_1$	Self-loops, N/A
2	$miRNAs_{same}, Diseases_{different}$	$M_1D_1 \rightarrow M_1D_2$	Present in the result set
3	$miRNAs_{different}, Diseases_{same}$	$M_1D_1 \rightarrow M_2D_1$	Present and used for analysis
4	$miRNAs_{different}, Diseases_{different}$	$M_1D_1 \rightarrow M_2D_2$	Present in the result set

Table 6.: Types of interactions in the network

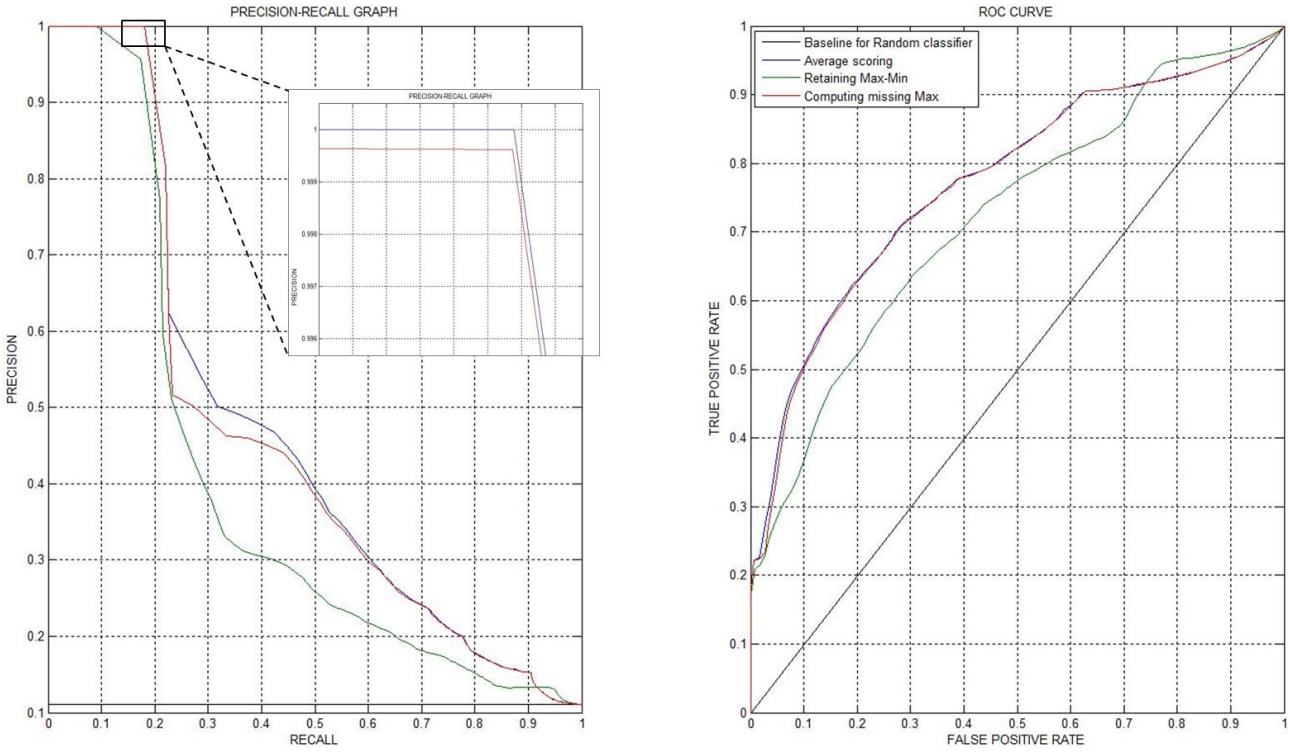


Fig. 25.: Precision-recall and ROC curves displaying the accuracy of the three methods. The figure demonstrates that the *Average scoring* (blue curve) method fared better than *Retaining Max-Min* (green curve) and *Computing Missing Max.* (red curve) methods. The inset image shows that the precision of *Average scoring* method slightly outperformed the *Computing Missing Max.* and was the best overall performer.

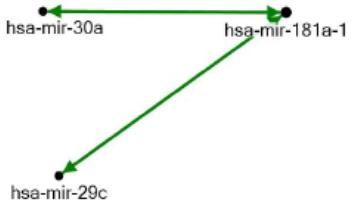
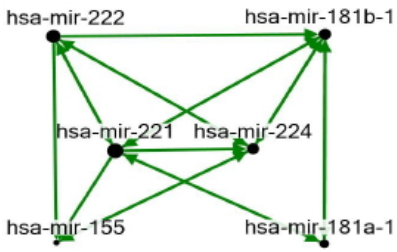
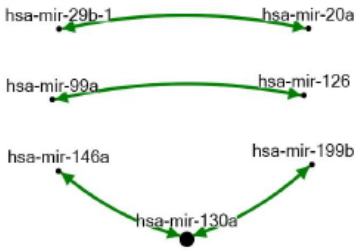
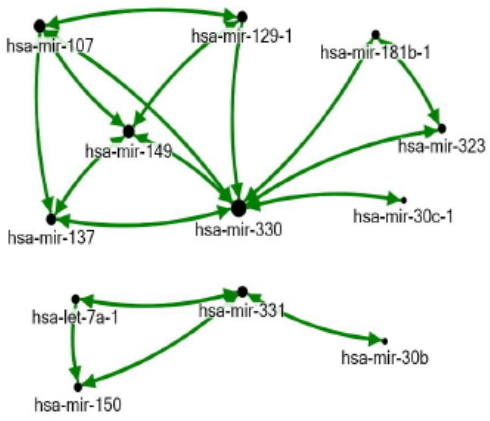
Category	Preserved miRNA-miRNA interaction	Critical miRNAs	PubMed IDs
Gastrointestinal Cancers		hsa-mir-30a hsa-mir-181a-1 hsa-mir-29c	Colorectal cancer: 18607389, 20480519, 22112324 Gastric cancer: 20022810, 21139804 Gastrointestinal cancer: 19030927, 24289824, 23950912 Gastroesophageal cancer: 19737949, 24289824 Esophageal cancer: 19737949, 20480519
Endocrine Cancers		hsa-mir-221, hsa-mir-222, hsa-mir-155, hsa-mir-224, hsa-mir-181a-1, hsa-mir-181b-1	HCC: 18223217, 21139804 Pancreatic cancer: 16192569, 16966691, 21139804, 24289824 Thyroid carcinoma, follicular: 18270258, 17205120 Thyroid carcinoma, papillary: 18270258, 17205120
Leukemia		hsa-mir-29b-1, hsa-mir-20a, hsa-mir-126, hsa-mir-130a, hsa-mir-99a, hsa-mir-146a, hsa-mir-199b	Hematological tumors: 16192569, 21139804 Leukemia, acute myeloid: 18337557, 21708028, 19602709 Leukemia, chronic lymphatic: 17934639, 20439436 Leukemia acute myelogenous: 18187662, 21708028, 19602709
Nerve Cancers		(A) hsa-mir-323, hsa-mir-129-1, hsa-mir-137, hsa-mir-330, hsa-mir-149, hsa-mir-107, hsa-mir-30c-1, hsa-mir-181b-1 (B) hsa-mir-30b, hsa-mir-331, hsa-mir-150, hsa-let-7a-1	(A)Medulloblastoma: 18973228, 24213470 Neuroblastoma: 17283129, 25238782 Glioblastoma somatic: 18577219, 17363563, 24213470 (B)Medulloblastoma: 18973228, 18756266, 25594007, 22623952 Neuroblastoma: 17283129, 24438171, 23220581 Glioblastoma somatic: 17363563, 26132860, 21139804, 26046581

Fig. 26.: Signature miRNA-miRNA interaction component identified in various cancer categories. The PubMed IDs citing the critical miRNAs with the disease from the *PhenomiR* database are in magenta while the PubMed IDs from the *PubMed Central* database are in blue.

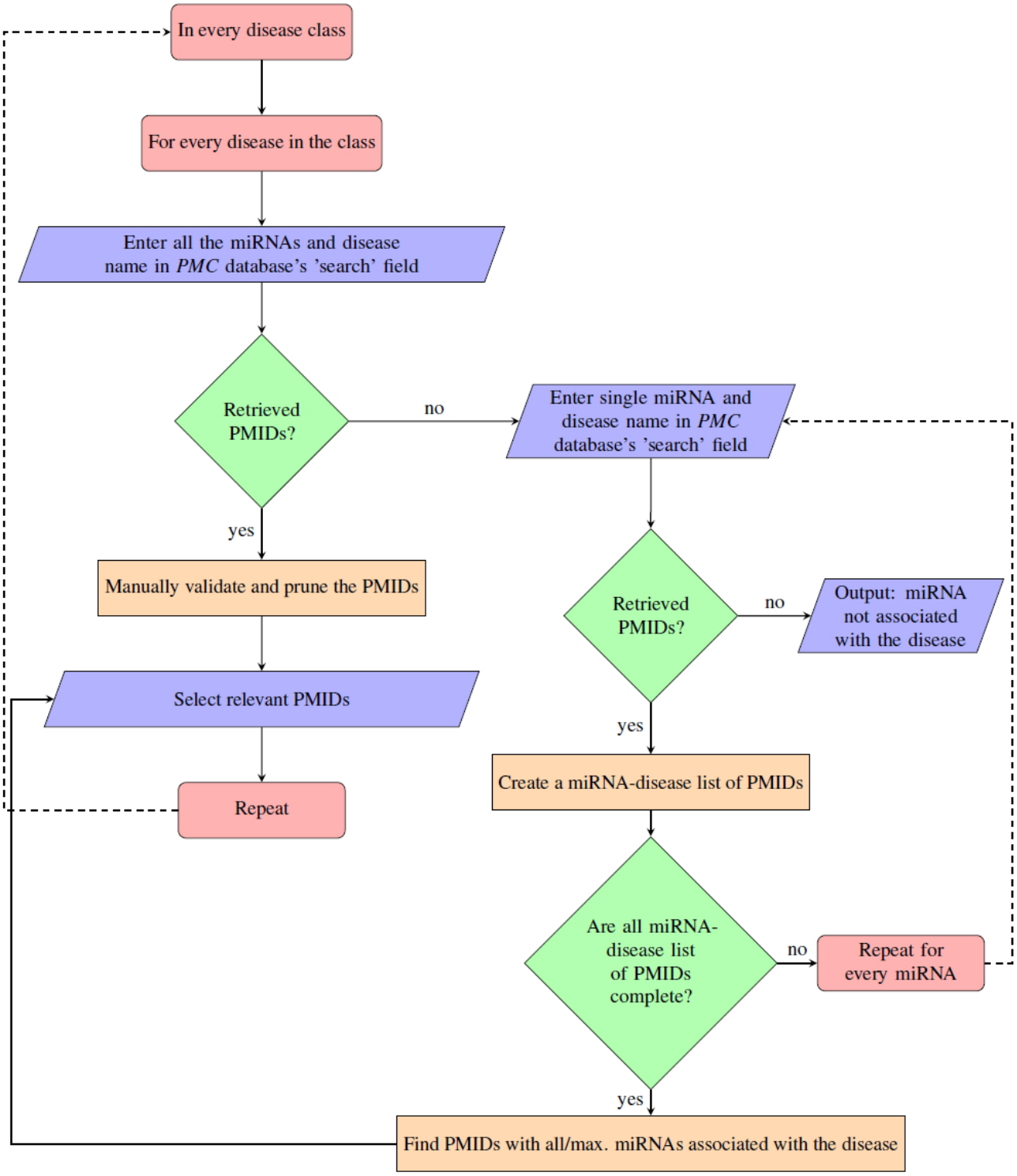


Fig. 27.: Flowchart of the work flow for manual literature search.

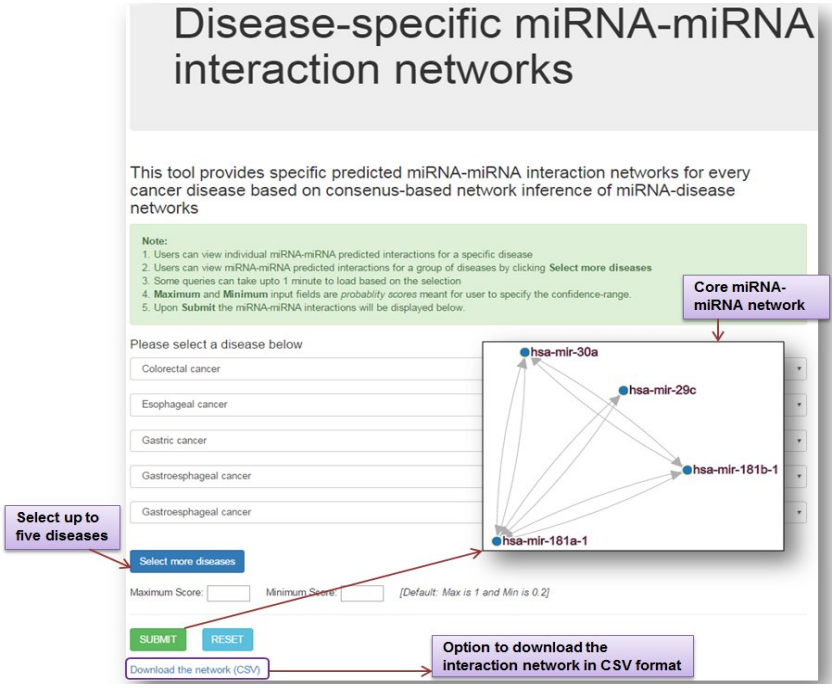


Fig. 28.: miRNA-miRNA interactions shown in *miRsig* for esophageal, gastroesophageal, gastrointestinal, gastric, and colorectal cancers.

CHAPTER 6

INFLUENCE DIFFUSION MODEL

In this study, we use information diffusion theory to quantify the influence diffusion in a miRNA-miRNA regulation network across multiple disease categories which we derived in our previous work (Chapter 5). Our proposed methodology determines the critical disease specific miRNAs which play a causal role in their signaling cascade and hence may regulate disease progression. We extensively validate our framework using existing computational tools from the literature. Furthermore, we implement our framework on a comprehensive miRNA expression data set for alcohol dependence and identify the causal miRNAs for alcohol-dependency in patients which were validated by the phase-shift in their expression scores towards the early stages of the disease. Finally, our computational framework for identifying causal miRNAs implicated in diseases is available as a free online tool for the greater scientific community.

6.1 Motivation

Although many studies have identified miRNAs associated with diseases, only a few of those have investigated the (signal) cascading influence/effect of miRNA (de)regulations onto other miRNAs or molecular participants. Despite the wide availability of data regarding a miRNA's direct/indirect effect on various biological processes, identifying or quantifying their influence remains a challenge. To the best of our knowledge, there has not been any model that simulates the time or an event-driven progression of miRNA (de)regulations leading up to a pathophysiological disorder. It is still unknown *how* (de)regulations of a miRNA impact a disease

progression and/or their repression. Understanding the progression of such miRNA-driven signaling cascade in the context of diseases is extremely crucial for identifying (i) the critical miRNAs (as potential biomarkers and *directors* of global expression patterns); and (ii) the key stages in the progression of a disease-state under the influence of miRNAs' expression.

In this work, we model the passage of miRNA-based influence propagation among other miRNAs as a network diffusion model. We use social behavioral/network principles to model a miRNA's cascading influence or *flow of information* in and among **d**isease-specific **m**iRNA **i**nteraction **n**etworks *(DMIN)* (elaborated in the *Methodology* section). Essentially, a *DMIN* is a (predicted) miRNA-miRNA interaction network pertaining to a specific disease. These networks often resemble the behavioral characteristics of a social network, such as homophily, [110] wherein participants tend to have positive ties with participants that are similar to themselves; this has already been evidenced in the case of miRNAs[20]. Hence, the application of social network algorithms is apt for modeling the progression of a miRNA's activity and its signal cascading effect in the context of a disease-state. We explore the property of *information diffusion* through miRNAs which is a crucial characteristic of a *DMIN* network and study the aspect of *information flow* in *DMIN*s. Consider a network of miRNA nodes as shown in Figure 29. At time point T_1, only *miRNA-1* is activated (in green color). At time point T_2, *miRNA-1* attempts to activate its neighbors, *miRNA-2* and *miRNA-3*. While, *miRNA-2* is not activated (shown by a red arrow), *miRNA-1* successfully activates *miRNA-3* (shown by a green arrow). At time point T_3, *miRNA-3* tries to activate *miRNA-4* and *miRNA-5*, out of which only *miRNA-5* gets activated (shown by a green arrow) while activation of *miRNA-4* is unsuccessful (shown by a red arrow). And at time point T_4, *miRNA-5* successfully activates *miRNA-4*. A particular disease-state is assumed to be highly probable once a required set of

crucial miRNA nodes in a network are activated. In this work, we refer to *activating/influencing* a miRNA as a function of time and analogous to affecting a miRNA's expression and activity. Note that miRNAs implicated in a particular disease may either be up- or down-regulated; our notion of activating/influencing a miRNA is abstract and encompasses both cases. In other words, the nodes in a *DMIN* sequentially activate/influence others where such activation pertains to a significant differential expression of a miRNA over its corresponding expression at control. In this work, we devise a modeling framework to identify the signaling cascade of miRNAs that have already been implicated in particular diseases; our framework can distinguish between causal miRNAs and the affected ones from the global pool of miRNAs that were implicated in a disease. We also present this framework in the form of an online web tool, *miRfluence* that can be readily used by the scientific community. Once the passage of influence between miRNAs is decoded based on the software tool presented here, it will motivate a wide variety of applications ranging from predicting disease progression, disease outcomes and designing drug therapeutics.

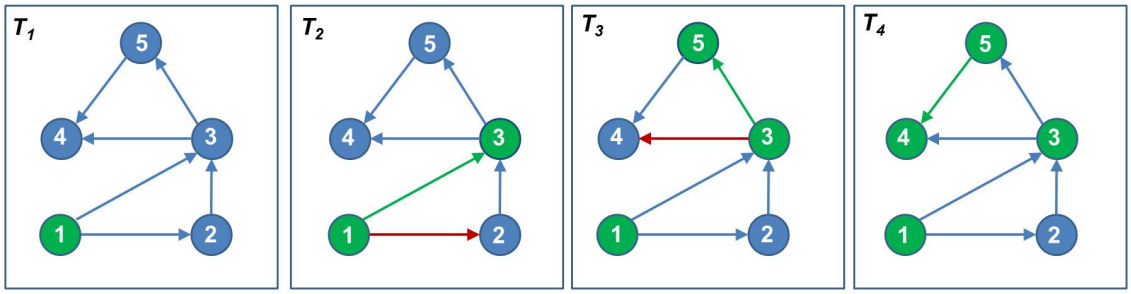


Fig. 29.: Cascading flow of influence in a *DMIN*

6.2 Background

The concept of information diffusion in a network has been widely deployed in the field of social network theory to study spread of ideas, rumors and product

adoption between the individuals in the network via the *word of mouth* effect [45–47]. There are essentially two fundamental models of information propagation in social networks - linear threshold *(LT)* and independent cascade *(IC)* model. Every other model proposed in the literature is a derivative of these canonical models. Although, this concept has been applied in the field of sociology to study the various behavioral phenomena, such as the spread of a new concept[48], it has also been extended to understand the dynamics of spreading of diseases[111–113]. However, understanding influence diffusion in a complex network of miRNAs has never been attempted before and is challenging due to the multi-level nature of interactions. In this work, considering that miRNAs of similar diseases tend to act cooperatively[20], we focus on the *social nature* of miRNAs related to a class of diseases. We deploy an information diffusion model, through which a miRNA's influence on its neighboring miRNAs is analyzed and quantified. Social influence can affect a range of behaviors in networks such as dissemination of information/influence, communication and in this case, even mutation. In both the LT and IC model, the nodes (i.e. the miRNAs) in the network can be in one of the two states - active or inactive. The activated nodes spread their influence by activating their neighboring inactive nodes based on a certain criteria or effect. Garnovetter et al. [49] proposed the LT model by applying a specific threshold in each of the nodes of the network. Therein, each node is activated only by its neighbor(s) depending upon the cumulative weight of the incoming edges to the node. The node becomes active when the cumulative sum of the weight of the incoming edges from an active neighboring node crosses its threshold value. Once activated, the node remains active and tries to activate its neighbor, thereby propagating its influence. On the contrary, the IC model uses edge probability to determine the information diffusion. In this model, an active node has a single opportunity to activate its neighbors. The edge weights represent the

activation probability or likelihood of information propagation in between two nodes. Hence, upon activation, an active neighbor is likely to choose a neighbor with the highest edge weight to activate next.

The miRNA-miRNA interaction network in *DMIN*s used in this study have probability scores as edge weights. These scores act as activation probabilities. Using the *IC* model, upon an activation of a certain miRNA, based on the edge weights between its neighbors, we can determine the next miRNA that is likely to be activated. In this context, activation implies having a causative effect on another miRNA's expression level. This effect may be direct (when a miRNA directly controls the expression of another one) or indirect (when such regulation can be due to intermediate genes/proteins that these miRNAs regulate). Following this pattern, the information flow or the spread of influence across the miRNAs can be detected. Hence, the pattern of influence across miRNAs in a disease can be identified and studied. Further, we integrate different *DMIN*s belonging to the same category profile, (e.g. 'gastrointestinal cancers') and detect the spread of influence among miRNA-miRNA interaction networks belonging to this profile. Subsequently, we determine the key miRNAs playing an influential role among all the diseases within a certain profile.

6.3 Methodology

6.3.1 Disease-specific miRNA-miRNA interaction networks *(DMIN)*

PhenomiR 2.0 database[68] is a manually curated comprehensive data set of differentially regulated miRNA expressions in diseases. It contains 632 database entries collated from 345 articles pertaining to 675 unique miRNAs and 145 diseases. The data curated in PhenomiR is not normalized and is available for download as is. An example of miRNA's foldchange values and their corresponding regulations in a

disease is shown in Figure 30(i), where miRNAs are denoted by $M_1 - M_5$ and disease is denoted by D_1. Nalluri et al, developed a consensus-based network inference pipeline from the PhenomiR dataset to predict key miRNA signatures (i.e., groups) across several categories of diseases in Chapter 5. To briefly summarize this work, they considered a pair of miRNA and disease as a single miRNA-disease (MD) entity (or node) which conceptually signifies a disease-specific miRNA. Therefore, the expression score of M_iD_j would mean the expression score of miRNA i in disease j. Next, they created a miRNA-disease expression matrix, in which the rows represented the various samples/studies and columns represented MD nodes. Next, they used six network inference algorithms on the expression matrix and a consensus-based aggregation approach to derive the probabilistic $MD - MD$ interaction network. From this $MD - MD$ interaction network, they further extracted **d**isease-specific **m**iRNA-miRNA **i**nteraction **n**etworks (*DMIN*)s and made them available in the tool, *miRsig*(mentioned in Chapter 5). *DMIN*s are directed graphs $G(V, E)$ where V is the set of miRNAs being regulated in a specific disease and E is the set of weighted edges between them denoting the probability of an interaction. We downloaded the *DMINs* from *miRsig* as is, and further developed an optimization-based methodology (detailed in the next section) to generate a modified *DMIN* which would serve as the input network for the influence diffusion based strategy (Figure 30(iv)). *miRsig* hosts *DMIN*s for 66 specific diseases. However, to pursue a defined cancer-specific analysis, only 17 *DMIN*s were considered. Based on their tissue-specificity, *DMIN*s were grouped into four categories, namely cancer of the gastrointestinal, endocrine, brain systems and leukemia resulting in four *DMIN*s corresponding to each category.

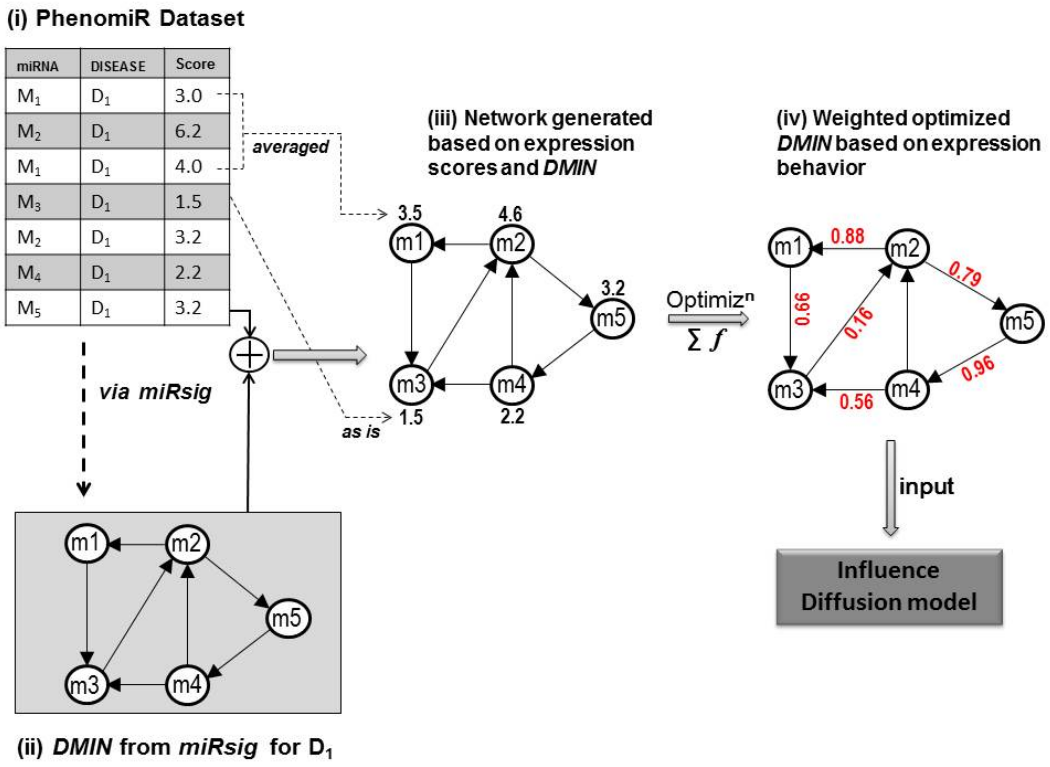


Fig. 30.: Overview of network generation via optimization of expression scores in a *DMIN*.

Network generation via optimization of expression scores

DMINs are directed miRNA-miRNA interaction networks with probability scores as edge weights. To have a network with highest confidence, we extracted *DMIN*s having edge weights of 0.90 score and above. Selecting a high cut-off of 0.9 on the edge scores is a standard practice in such reverse engineering algorithms to ensure confidence in the results. Such algorithms generally suffer from low accuracy due to the noisy expression datasets, non-linearity in the miRNA interactions as well as the high complexity of the inverse problem of inferring N^2 edges in a network of N nodes. Hence, it is customary to work with only the high-confidence edges signified by a 0.9 cut-off on the edge scores. Additionally, the very nature of the influence diffusion set-up works better for sparse graphs; for more dense graphs, most nodes in the network will end up having a high influence score for activating the entire network just because of the availability of more paths to destination nodes. To avoid this possibility, we chose a edge score cut-off of 0.9 in this paper. After deriving these *DMIN*s, we discarded the edge weights (Figure 30(ii)). We term this network as $DMIN_{HC}$ (*DMIN* of **h**igh-**c**onfidence). Although, $DMIN_{HC}$ captures the miRNA-miRNA interaction topology, it does not take into consideration the expression scores of the individual miRNAs within their corresponding diseases (Figure 30(i)). Expression scores are a vital part of the miRNA-disease regulatory mechanism. Hence, we append $DMIN_{HC}$ with expression scores for every node (i.e., miRNA), as shown in Figure 30(iii) resulting in our final network, $DMIN_{HCE}$(*DMIN* of **h**igh-**c**onfidence and **e**xpression score). The expression scores of miRNAs were converted to their log_2 scores before being incorporated. While incorporating miRNA's expression scores into *DMIN*, some miRNAs had multiple expression scores for a particular disease (e.g., row #1 and #3 in Figure 30(i)). In such instances, these multiple scores were

averaged to get the best possible estimated expression score to be incorporated into the *DMIN*(e.g., node M_1 in Figure 30(iii)). Nalluri et. al., demonstrated that *averaging* of the multiple expression scores yielded the best estimate for *DMINs* (*Average Scoring*, under Methods), Chapter 5.

It is important to note that, in our network-building methodology of $DMIN_{HCE}$, we justify the underlying biological implications of a miRNA's regulatory behavior. We assume that expression changes in a particular miRNA will have consequential effect on another miRNA's expression behavior. Hence, if a miRNA has a very high expression score (i.e. degree of fold-change) and is connected to its neighboring miRNAs then it would have a corresponding degree of influence or propagating effect on its neighboring miRNAs. Hence, in order to build a network model which is as close to the underlying biological activity, we design the following optimization-based strategy which provides us with $DMIN_{HCE}$ with edge-weights, i.e., a weighted $DMIN_{HCE}$ (see Figure 30(iv)). These edge-weights would quantify influence of one miRNA onto another, thereby modeling the behavior of miRNA's regulatory activity based on their expression scores.

Optimization formulation for generating edge weights

In order to derive the edge weights for $DMIN_{HCE}$, the following assumptions were postulated.

1. The direction implies regulatory influence.

2. Each miRNA's expression score is a cumulative result of its neighboring miRNAs' expression scores. Hence, the cumulative sum of incoming edge-weights would equal to the expression score of the miRNA. This is denoted by the *Incoming* constraint in the optimization function.

3. Each miRNA's outgoing edge-weights would not exceed its expression score. A miRNA's expression score corresponds to its outgoing edge-activity implying that the consequential effect a miRNA has on its neighboring miRNAs is directly correlated to its expression score. However, in this case we introduce a *slack* quantity to make the model more relaxed and feasible for solutions. Without the *slack* variable, the model becomes too restrictive and would not yield any solutions. This is denoted by the *Outgoing* constraint in the optimization function.

The formulation is as follows,

Objective: To achieve the most optimal regulatory network flow (i.e., edge weights) characterized by expression scores of each node. This is achieved by obtaining minimum slack (denoted by s_i) throughout the network; subject to constraints that (i) the cumulative sum of products of incoming edge-weights and corresponding expression scores of parent nodes would equal the expression score of the target node and (ii) sum of every node's outgoing edge-weights can exceed its expression score within a *slack* amount.

Variables: Let $X_{i,j}$ be the flow of influence from node i to node j, where $i, j \in n$, and n is the total number of nodes in the network, e be the fold-change expression of a node, and s be the *slack* quantity for $\forall i, j$

Optimization function:

$$\text{minimize} \quad \sum_{i=1}^{n} \mid s_i \mid$$

$$\text{subject to constraints:} \quad Incoming_{expression-flow} \sum_{j=1}^{n} e_j * X_{j,i} = e_i (i = 1, 2, ..., n)$$

$$Outgoing_{expression-flow} \sum_{j=1}^{n} e_i * X_{i,j} + s_i = e_i (i = 1, 2, ..., n)$$

where

$$0 \leq X_{i,j}; i \neq j \tag{6.1}$$

The above methodology provided an optimally-weighted $DMIN_{HCE}$ (Figure 30(iv)) which was used as the input network for the subsequent influence diffusion algorithm.

The goal of the optimization step is to derive as good an input for the subsequent analysis for influence diffusion based on the expression behavior of miRNAs. Note that ideally for the influence diffusion analysis, the edges should signify the influence of the source node onto the target node. In terms of chemical kinetics of the $A \rightarrow B$ edge, such influence is determined by [concentration of the source node A]×[rate constant]; since such rate constants of the miRNA interaction network are unknown (and very difficult to validate experimentally as they comprise indirect interactions of possibly multiple components), we simply considered the influence of an edge to be governed by the concentration of the source node exclusively. Also, considering the steady-state concentrations of the source nodes only signify the equilibrium edge weights and hence the static influence of the source onto the target; this does not capture the time varying influence on the edges as the source node concentration should ideally vary with time.

In addition to the optimization-based network generation method, we also implemented another network generation method - *'Rescoring all edges to constant weight'*, wherein we assign a constant weight on all the edges of the network. The goal of the more simplistic constant edge weights is to further disregard the steady state source node concentrations and assign equal weightage to all edges in terms of their influence. This formulation can only identify the topological pressure points and should be less accurate. However, due to limitations on the availability of such detailed time-series datasets on miRNA expression levels in specific diseases, it is currently not possible to quantitatively show the difference in accuracy of identifying the influential miRNAs from the two approaches. Perhaps, the common miRNAs that show up to be influential from both approaches will be a better option to consider.

Influence Diffusion analysis

Upon deriving the weighted $DMIN_{HCE}$ for 17 diseases and four disease categories, we implemented the influence diffusion algorithms to derive a list of miRNAs ranked according to their highest influence in a disease category (see Section *Compute Influence* and *Algorithm 1*). This algorithm was implemented using the influence maximization code freely distributed[114].

In a $DMIN_{HCE}$ of a disease category, there may be multiple occurrences of the same miRNA-miRNA interacting edge due to its presence among several diseases of the category. Here, two approaches are further adopted to calculate their single edge prediction score. As seen in Figure 31(a), two weighted $DMIN_{HCE}$s belonging to disease D_1 and D_2 are under the same disease category. The edge m_1-m_2 is present in both networks with different edge-weights. To address these scenarios, we devised the following two approaches.

i) *Logical AND/ Intersection* operation

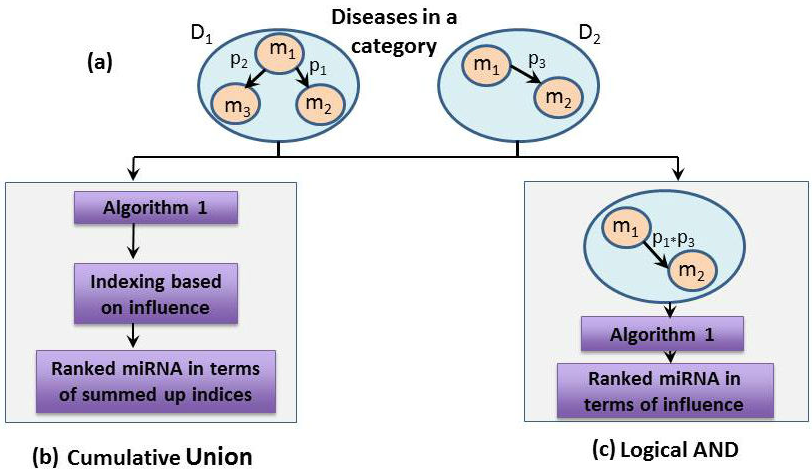


Fig. 31.: Overview of the work flow of the methodology. Consider two weighted $DMIN_{HCE}$s belonging to disease D_1 and D_2 which are under the same disease category. The edge $m_1 - m_2$ is present in both the networks. In the final updated network, the edge weight of $m_1 - m_2$ is recalculated accordingly using the *Logical AND* operation and upon this updated network, the *Compute Influence* algorithm is implemented.

Under this operation, only the edges which were present in all the diseases of a category were retained in the final disease category network. The edge weights for these common miRNA-miRNA interaction edges were calculated by the following formula,

$$P_{new} = P_1 \times P_2 \times ... \times P_n \tag{6.2}$$

where P_1, P_2, and P_n are prediction scores of the same edge in individual disease networks.

This operation was implemented on the following four categories consisting of the subsequent diseases:

Gastrointestinal category: esophageal carcinoma, gastroesophageal carcinoma, gastrointestinal cancer, gastric cancer, colorectal cancer.

Leukemia category: hematological tumors, acute myeloid leukemia (AML), susceptibility to chronic lymphatic leukemia, acute myelogenous leukemia.

Endocrine category: pancreatic cancer, hepatocellular carcinoma (HCC), thyroid carcinoma (follicular), thyroid carcinoma (papillary).

Brain systems: neuroblastoma, medulloblastoma, glioblastoma.

ii) *Cumulative Union*:

Under this approach, firstly, for every weighted $DMIN_{HCE}$ in the category, each miRNA's coverage was determined using *Algorithm 1.* The coverage value of each miRNA was mapped into a coverage-percentage (e.g., a node having coverage-percentage score of 70 would imply its influence over 70% of the nodes in the network). Next, for the miRNAs which were repeated in multiple diseases within the category - their coverage-percentages were averaged. Finally, the miRNAs are ranked as per their coverage-percentage in the disease category. An explanation of the coverage compu-

tation algorithm is provided in the next section.

Compute Influence (coverage)

This algorithm (i.e., *Algorithm 1*) is based off of the *IC* model of *information diffusion*. Let $COV(u)$ denote the coverage/influence of a miRNA node u in the network. Upon the execution of the algorithm, all miRNAs are ranked as per their highest coverage/influence. The coverage of each node has been calculated after 10000 monte carlo simulation cycles to achieve the optimal value of coverage.

The algorithmic details of computing the COV function are described in the theory of *Independent Cascade* model stated in Kempe et al's work[115]; however the following is the summary of its working.

1. Select a node in the network, e.g. consider node 1 in Figure 32(T_1).

2. Along its every outgoing edge, perform a biased coin toss, where bias is the edge-probability. In Figure 32(T_2), this operation is performed along edges 1→2 and 1→3 having edge weights 0.5 and 0.9, respectively.

3. If the coin toss operation is successful, then activate the node, and perform step (2) on the newly activated node. In Figure 32(T_3), node 2 is not activated (denoted by a red edge, 1→2) while node 3 is activated (denoted by a green node 3 and edge 1→3). Next, a biased coin toss is performed on node 3 along its edges 3→4 and 3→5 which results in activation of node 5 and a failed activation of node 4. Subsequently, step (2) is performed on node 5 which results in activation of node 4.

4. Stop when there are no more new activations possible. In Figure 32(T_4), no more new activations are possible. Hence, the nodes which can be influenced by node 1 are nodes 3, 5, and 4. Coverage score of node 1 is three.

5. Perform steps (2-4) for the initially selected node (i.e., node 1 in Figure 32) 10,000 times and finally, average the coverage scores.

6. Repeat steps (1-5) for next node.

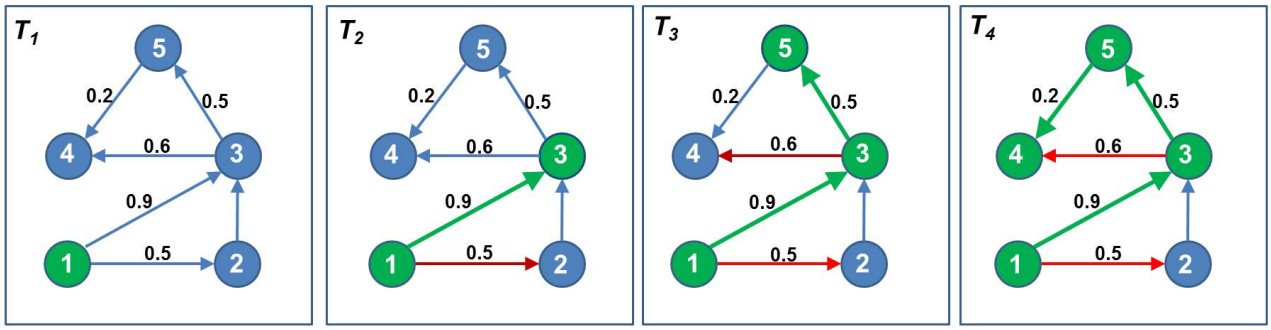


Fig. 32.: Computation of coverage of influence for node 1. Node 1 activates node 3, node 5 and node 4 based on a series of biased coin-toss operations along its edges

```
//Algorithm: Computing coverage of every node in the network
//No. of Monte Carlo cycles
N=10000
  //For every miRNA node
for u = 1 to V
  I = 0;
    for i = 1 to N
          //Calculate coverage of Node u
          I(u)+= COV(u)
        end for
   //Normalize the result
  I(u) = I(u)/N
end for
```

```
//Sort the miRNA nodes as per their influence
Sort(V, I)

//Algorithm runtime = O(vRm);
//-----------------------------
v = number of nodes
R = number of repeated simulations
m = number of edges
```

Results

The above methodology was implemented on $DMIN_{HCE}$s of four disease categories and on individual diseases as well. However, to maintain the emphasis on pan-cancer diseases, we discuss the results of this methodology on the aforementioned disease categories. The results of individual diseases are available and can be downloaded for study and research from the tool *miRfluence*. We implemented the two approaches, i.e., *Intersection/Logical AND* and *Cumulative Union* for the $DMIN_{HCE}$s of these four disease categories, and the results are labeled under *Influence Maximization* in Table 7. Under the *Cumulative Union* approach, since all the miRNAs (belonging to a category) are ranked as per their coverage percentage, we have selected top 10 miRNAs to be displayed as most influential miRNAs. The results obtained were compared with two other approaches - *miRsig*(Chapter 5) and tool for annotations of miRNAs *(TAM)* [116]. *miRsig* uses a consensus-based network inference pipeline to predict the crucial miRNAs among the disease categories. The TAM method uses a prediction model to identify novel miRNA interactions and the most likely diseases to be affected (noted with *p-values*) for the input set of miRNAs.

It is also important to note that there are hardly any tools which predict/determine a list of crucial miRNAs *based on an input set* of diseases. The availability of tools which predict a set of diseases *based on an input set of miRNAs* are also scarce (like tool for annotations of miRNAs(*TAM*)). Many tools provide individual *miRNA-disease* associations and prediction scores but not *set-onto-set* analysis. These factors make *one-on-one* comparison of the proposed methodology very challenging. Hence, we have used the only tools that are available for comparison. The results are presented in Table 7.

1. *Endocrine cancers* (see row *Endocrine cancers* in Table 7)

   As per *Logical AND/ Intersection* approach, the miRNAs hsa-mir-181b-1, hsa-mir-181a-1, hsa-mir-224, hsa-mir-221 and hsa-mir-222 are key influential miRNAs and they were also predicted as crucial miRNAs as per the tool, *miRsig*. These same miRNAs are also present in the list of top ten miRNAs under the *Cumulative Union* approach. As per the tool *TAM*, all the diseases of this category, i.e., *thyroid neoplasms*, *pancreatic cancer* and *HCC* are very likely to be associated with the aforementioned list of miRNAs. The reported PubMed IDs report the occurrence/expression of all the resultant miRNAs within the same PubMed ID.

2. *Leukemia* (see row *Leukemia cancers* in Table 7)

   Under this category, all the miRNAs determined by the *Logical AND/ Intersection* were identified as crucial by the tool *miRsig*. Seven miRNAs predicted by the *Cumulative Union* approach are confirmed by *miRsig*, as well. Among the diseases, acute myeloid leukemia (AML), chronic lymphatic leukemia (CLL) and hematological disorders, were determined as most likely diseases as per *TAM*. The reported PubMed IDs report the occurrence/expression of all the resultant

Category	Methods					PubMed IDs
	Influence Maximization		miRsig	TAM (disease: p-value)		
	Intersection	Cumulative		Intersection	Cumulative	
Endocrine cancers	hsa-miR-181b-1	hsa-miR-224	hsa-miR-221	thyroid neoplasm: 2.34e-04	thyroid neoplasm: 2.56e-03	18270258
- hepatocellular carcinoma (HCC)	hsa-miR-181a-1	hsa-miR-155	hsa-miR-222	pancreatic: 4.61e-03	pancreatic: 5.76e-04	21139804, 24289824, 16966691
- pancreatic cancer	hsa-miR-224	hsa-miR-222	hsa-miR-155			
- thyroid carcinoma, follicular	hsa-miR-221	hsa-miR-181a-1	hsa-miR-224			
- thyroid carcinoma, papillary	hsa-miR-222	hsa-miR-181b-1	hsa-miR-181a-1			
		hsa-miR-221	hsa-miR-181b-1			
		hsa-miR-187				
		hsa-miR-31				
		hsa-miR-205				
		hsa-miR-181c				
Leukemia cancers	hsa-miR-130a	hsa-miR-126	hsa-miR-29b-1	AML: 1.12e-02	AML: 3.48e03	18337557, 21708028, 19602709
- hematological tumors	hsa-miR-199b	hsa-miR-130a	hsa-miR-20a	CLL: 1.92e-02	CLL: 8.59e-03	17934639, 20439436
- acute myeloid leukemia (AML)	hsa-miR-29b-1	hsa-miR-20a	hsa-miR-126	hematological: 6.53e-03	hematological: 0.127	16192569, 21139804
- chronic lymphatic leukemia (CLL)	hsa-miR-146a	hsa-miR-29b-1	hsa-miR-130a			
- acute myelogenous leukemia	hsa-miR_20a	hsa-miR-99a	hsa-miR-99a			
		hsa-miR-199b	hsa-miR-146a			
		hsa-miR-106a	hsa-miR-199b			
		hsa-miR-146a				
		hsa-miR-222				
		hsa-miR_155				
Gastrointestinal cancers	hsa-miR-181a-1	hsa-miR-29c	hsa-miR-30a	None	Colorectal cancer: 3.06e-02	18607389, 20480519, 22112324
- esophageal	hsa-miR-30a	hsa-miR-181a-1	hsa-miR-181a-1			
- gastroesophageal	hsa-miR-29c	hsa-miR-30a	hsa-miR-29c			
- gastrointestinal		hsa-miR-181b-1				
- gastric		hsa-miR-195				
- colorectal		hsa-miR-221				
		hsa-miR-21				
		hsa-miR-210				
		hsa-miR-99a				
		hsa-miR-126				
Brain systems	hsa-miR-330	hsa-miR-187	hsa-miR-323	None	Glioblastoma: 0.09	17363563, 18577219, 24213470
- neuroblastoma	hsa-miR-149	hsa-miR-181b-1	hsa-miR-129-1		Medulloblastoma: 0.29	18973228, 24213470, 18756266
- medulloblastoma	hsa-miR-331	hsa-miR-137	hsa-miR-137			
- glioblastoma	hsa-miR-107	hsa-let-7a-1	hsa-miR-330			
	hsa-miR-129-1	hsa-miR-150	hsa-miR-149			
		hsa-miR-190	hsa-miR-107			
		hsa-miR-323	hsa-miR-30c-1			
		hsa-miR-107	hsa-miR-181b-1			
		hsa-miR-149	hsa-miR-30b			
		hsa-miR-331	hsa-miR-331			
			hsa-miR-150			
			hsa-let-7a-1			

Table 7.: Results of influence maximization methods - *Intersection* and *Cumulative* compared to tools - *miRsig* and *TAM*

miRNAs within the same PubMed ID.

3. *Gastrointestinal cancers* (see row *Gastrointestinal cancers* in Table 7)

   The miRNAs predicted as influential (by *Logical AND/ Intersection*) under this category were also predicted to be critical miRNAs by the tool, *miRsig*. The top ten miRNAs predicted by the *Cumulative Union* approach had three of them confirmed by *miRsig* as well. In the gastrointestinal category, *colorectal cancer* is listed in TAM with a *p-value* of 2.03e-3. The reported PubMed IDs report the occurrence/expression of all the resultant miRNAs within the same PubMed ID.

4. *Brain systems* (see row *Brain systems* in Table 7)

   Under this category, all the miRNAs determined by the *Logical AND/Intersection* appraoch were predicted to be crucial by *miRsig*. Eight out of then reported miRNAs under the *Cumulative Union* approach were corroborated by *miRsig*. *TAM*'s prediction scores for the two diseases (*glioblastoma* and *medulloblastoma*) are not in the confidence margin. However, the reported PubMed IDs report the occurrence/expression of all the resultant miRNAs within the same PubMed ID.

Case Study and Proof-of-concept

Our proposed methodology is able to identify influential miRNAs in disease-specific networks as demonstrated in the previous section. Furthermore, since the dynamics of miRNA-mediated regulations are similar in biological networks, this methodology has broad applications ranging from networks pertaining to cancers to other pathophysiological conditions, as well. We further demonstrate the application of our proposed methodology on a miRNA expression data set generated from a

postmortem brain tissue from patients diagnosed with alcohol dependence (AD).

Tissues for 18 AD patients and matched controls were obtained from a larger sample of 41 AD cases and 41 controls. The postmortem brain sample was received from the Australian Brain Donor Program, New South Wales Tissue Resource Centre, at the University of Sydney, (http://sydney.edu.au/medicine/pathology/trc/). The demographic characteristics of the sample are described elsewhere[117].

The miRNA expression data were generated using the Affymetrix GeneChip miRNA 3.0 array and normalized using log_2 transformation, followed by quantile normalization, and median-polish probe-set summarization. The final miRNA expression data had 1733 miRNAs and 35 sample tissues (AD- 18, control-17). This expression data is provided in the Supplementary material.

The implementation of our proposed methodology on this data set consisted of the following steps:

1. Construction of a probabilistic miRNA-miRNA interaction network from the miRNA expression matrix based on the *miRsig* pipeline(Chapter 5). This network had 1733 miRNAs.

2. From this network, in order to generate a high-fidelity network, we consider only the edges which have a probability score of 0.9 and above.

3. Next, we determine 115 AD-related miRNAs. These miRNAs were derived from a brief literature survey mentioned in Ponomarev's work[118] which included miRNAs identified by Sathyan et. al.(2007)[119], Wang et. al. (2011)[120], Yadav et. al.(2011)[121], Lewohl et. al.(2011)[122] and Nunez et. al.(2012)[123]. We extract a sub-network consisting of 115 miRNAs and the edges among them from the larger network of 1733 miRNAs. This sub-network is a $DMIN$ for the disease condition - *alcoholic dependency*.

4. Next, we choose the option of re-scoring the edges of this network with a fixed edge score of 0.01. Note that since this AD dataset involves multiple expression values of each miRNA pertaining to each sample (18 AD and 17 control samples), it is not possible to directly use the optimization formulation for generating edge weights as discussed before; averaging the miRNA expression scores across both control and AD samples will not work here as the AD samples showed significantly different expression levels based on the number of years of alcohol consumption of the patients. Additionally, we did not directly use the edge probabilities from the consensus methodology for generating the miRNA network as such probabilities quantify the *feasibility* of an edge between two miRNAs and not the actual *influence* one miRNA has on the other one. Also, the $DMIN_{HC}$ is a highly dense and inter-connected network and hence having higher scores of edge weights will cause all the miRNAs to activate its neighbors, thereby labeling all the miRNAs as influential. Moreover, since these edge weights model *regulatory influence and flux phenomena*, lower values are more close to actual biological notion of flux dynamics; note that our goal here is to really understand the topological pressure points in this miRNA interaction network with a constant edge weight of 0.01 on all edges using the influence diffusion model. We tried other (constant) low edge scores as well and the rankings of the miRNAs based on coverage were very similar to the ones obtained here.

5. This *DMIN* is provided as input to the *influence diffusion* model (see *Algorithm*).

The result of the above implementation is a ranked list of miRNAs along with their coverage scores. Here, the coverage score implies the number of miRNA nodes that can be activated. For our further comparative analysis we consider the top five

Category	miRNAs
Top 5 miRs with highest influence	hsa-miR-376c
	hsa-miR-27a
	hsa-miR-30e
	hsa-miR-194
	hsa-miR-9
Bottom 5 miRs with least influence	hsa-miR-196a*
	hsa-miR-606
	hsa-miR-7b*
	hsa-miR-302b*
	hsa-miR-302c*

Table 8.: miRNAs with the highest and lowest coverage scores after the implementation of *Algorithm 1*

miRNAs with highest coverage and bottom five miRNAs with lowest coverage scores. This ranked list of miRNAs is provided in the Supplementary material.

Comparative analysis

The influence diffusion phenomena within a miRNA-miRNA interaction network is a time/event-driven progression, characterized by a series of (un)successful activations of miRNA nodes, as explained in Figure 32. However, the miRNA expression data set of alcohol-dependent patients used in this case study is not a time-series data set. The samples record the *Total years of drinking* alcohol for each patient. The *Total years of drinking* for these 18 samples are - 14, 20, 20, 24, 26, 27, 28, 29, 31, 31, 31, 32, 34, 36, 37, 39, 48 and 48. For the purposes of our modeling and in order

to introduce an element of time/event-driven series of progression to the miRNA-miRNA interaction network, we presume these individual samples as time points and observe the expression profiles of the miRNAs in Table 8, across these samples. Our hypothesis is that, these influential miRNAs would have undergone a phase-shift or a distinct change in their expression trend in the beginning stages of the time-points so as to signify an activation moment. This change may correspond to the triggering of the influence diffusion cascade process by the influential miRNAs. On the contrary, the least influential miRNAs would exhibit a similar trend to their control trend with possible phase-shifts occurring only in the later time points.

We plot the expression scores of these miRNAs (listed in Table 8) against the samples with number of *Total years of drinking*. For same sample time points (such as 20, 31 and 48), we average the expression scores of the miRNAs across AD and *control* samples in order to derive a single time point expression score. The expression trends (*AD* vs *control*) of top five miRNAs are displayed in Figure 33- a, c, e, g, i *(left side)* and those of bottom five miRNAs are displayed in Figure 33- b, d, f, h, j *(right side)*. The expression trends demonstrate that the top 5 miRNAs in AD-samples underwent a phase-shift in the beginning stages (especially around *year* 26) of the time-line when compared to their *control* trend, signifying a triggering of influence diffusion activity within the network. Conversely, the expression trends of the bottom 5 miRNAs in AD-samples align quite well with their *control* trend exhibiting slight fluctuations at later time points. The results corroborate our earlier stated hypothesis. The expression trends of the top 5 miRNAs also demonstrate that the miRNAs in the AD-samples were operating at a higher expression score from the start, signifying that they were already *activated* and were on an *ON* state. In order to better quantify the differences in their expression trends before and after the phase-shift with respect to the control, we conducted differential expression analysis of these miRNAs using

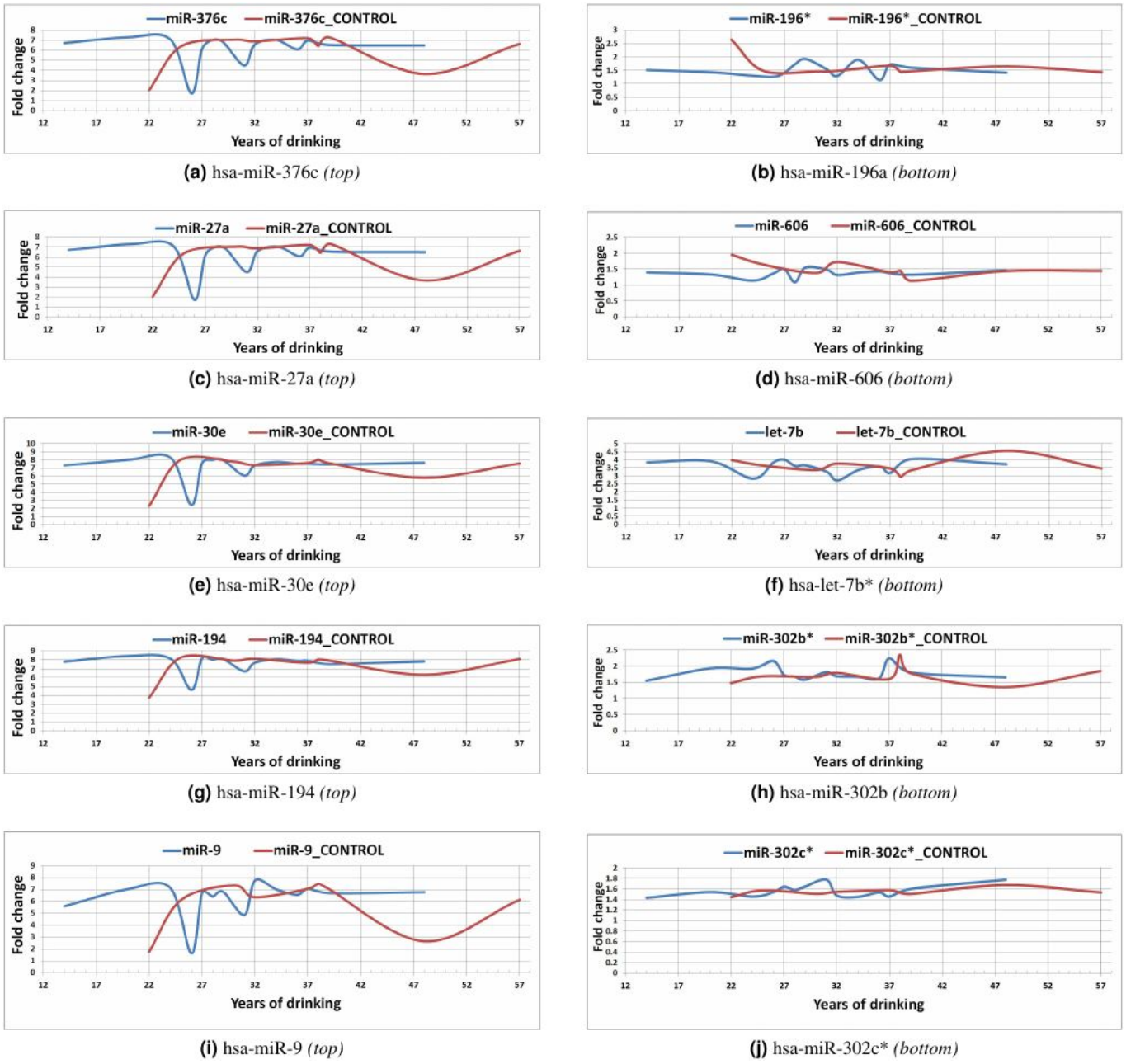


Fig. 33.: Trendlines of expression scores (AD vs control samples) of miRNAs with highest influence (a, c, e, g, i) and of miRNAs with lowest influence (b, d, f, h, j) across sample time points

miRNA	**Differential expression (p-values)**	
	Pre-phase shift	***Post-phase shift***
hsa-miR-376c	9.97e-07	0.539
hsa-miR-27a	5.13e-08	0.573
hsa-miR-30e	2.93e-07	0.503
hsa-miR-194	4.62e-06	0.523
hsa-miR-9	1.34e-06	0.829

Table 9.: Significance of differential expression of top 5 miRNAs *before* and *after* undergoing a phase-shift. Pre-phase shift p-values indicate there was a significant difference in the expression of their trends while post-phase shift p-values indicate that the expression trends did not differ significantly, as noted from Figure 33.

the *limma* package [124] from R-Bioconductor. We performed this analysis across two groups of data set: pre-phase shift and post-phase shift. For the purposes of this analysis, we chose the time-point of *year 26*, as the dividing time-point. The differences in the significance of the expression trends are shown in Table 9. Table 9 demonstrates that the difference in the expression trends of these miRNAs were very significant during the pre-phase shift period with respect to control in comparison to the post-phase shift period. This further emphasizes our hypothesis that the miRNAs underwent a phase-shift signifying the triggering of the influence diffusion cascade process towards the beginning stages of AD.

A point to note is that conventional differential expression analysis along with *in-vivo* strategies implicated all 115 miRNAs considered here to play a role in alcohol dependence; so the bottom 5 miRNAs from our list were also implicated in alcohol

dependence, albeit we argue that they were more of an *effect* of the signaling cascade, while the top 5 miRNAs exhibit more of a *causal* role.

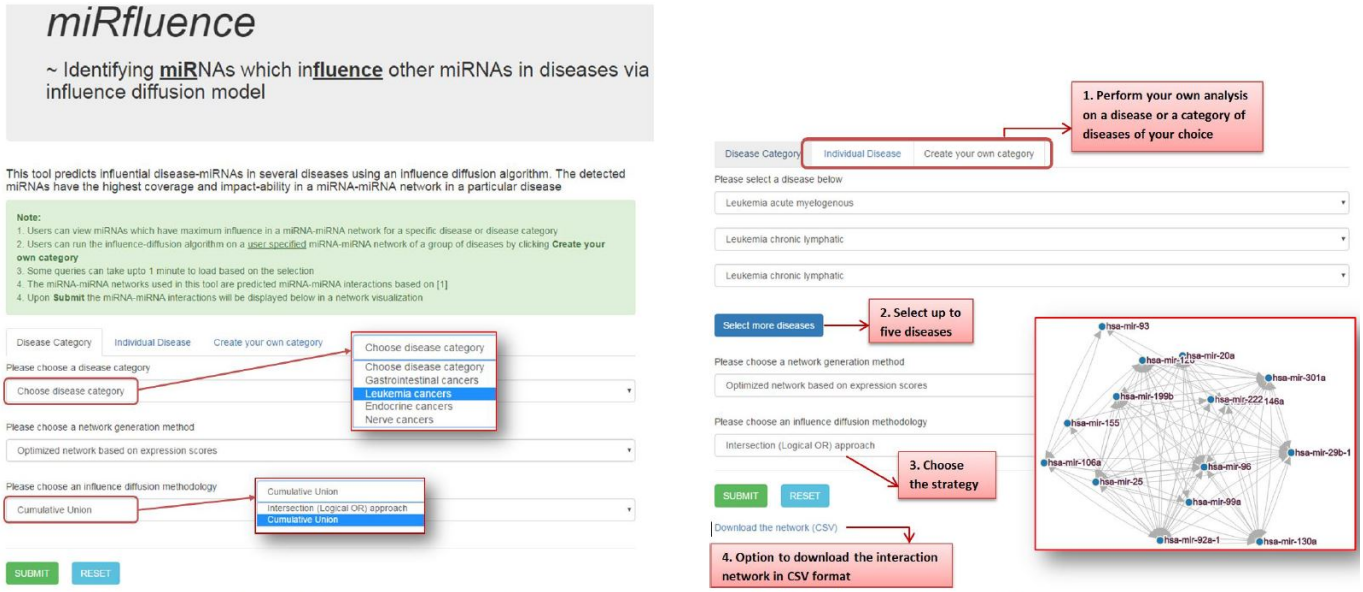
The expression trends displayed in Figure 33 of the miRNAs (listed in Table 8) demonstrate that the *influence diffusion* based methodology is able to identify top influential miRNAs playing a causal role in the miRNA interaction network, corroborated quantitatively by the expression trends of these miRNAs across the samples.

***miRfluence* - an influence diffusion implementation framework**

In order for researchers to implement the proposed influence diffusion methodology on various disease-specific miRNA-miRNA networks or on miRNA networks pertaining to diseases of their interest, we have developed *miRfluence*, an online platform. Using this platform, users can view the influential miRNAs in the miRNA-miRNA networks of existing categories and diseases (Figure 34-a). Users can also implement this methodology on a miRNA interaction network pertaining to any disease of their choice or can also create their own disease category with a combination of up to five diseases (Figure 34-b) from the existing set. Users can view the miRNAs and the topological placement of these miRNAs in the disease network. *miRfluence* also includes two options for identifying the edge weights of the miRNA interaction networks under the *Network generation method* option; these are the (i) optimized network based on expression scores and (ii) rescoring to 0.01 for all the edges considered above the 0.9 cut-off. Users can also choose the two types of influence diffusion implementations described in this work, namely *Logical AND/Intersection* and *Cumulative Union* approach. This tool will help researchers compare/contrast the influence of various miRNAs in similar/contrasting diseases and provide them an insight into the working and grouping of communities of miRNAs in an interactive visualization

making comprehension intuitive. The miRNA-miRNA interaction networks can also be downloaded in CSV format which can be easily imported into various network analysis tools for further study and analysis.

miRfluence is freely available for research purposes at http://bnet.egr.vcu.edu/mirfluence and has been developed using MySQL as the back-end database and Javascript, PHP, d3.js, AJAX and HTML/CSS for front-end design and visualization.



(a) Implement influence diffusion on existing data sets

(b) Implement influence diffusion on disease/s of your choice

Fig. 34.: *miRfluence* - an influence diffusion implementation framework

In this work, we have implemented the *information diffusion* concept from social networks to identify a crucial set of ranked miRNAs playing an influential role in diseases of a specific profile. Using this methodology, we were able to detect key influential miRNAs in the categories of *Gastrointestinal cancers*, *Leukemia*, *Brain cancers* and *Endocrine cancers*. These results were observed to be significant and were further validated by *miRsig* and *TAM* based analysis. For further validation, we used a miRNA expression data set of patients with alcohol-dependency; our top-ranked miRNAs indeed showed up to have possible causal effects in the miRNA

signaling cascade by showing phase-shifts in their expression towards the beginning stages of alcohol consumption in patients.

In our analysis, both the approaches used, i.e., *Logical AND/Intersection* and *Cumulative Union* produced similar results. Among the four categories, with the exception of *Brain cancers* all the miRNAs listed under the *Logical AND/Intersection* approach were included in the top ten ranks of the *Cumulative Union* approach which listed the miRNAs based on highest coverage scores. Hence, a more clear consensus as to which method fared better would emerge by testing these approaches on more comprehensive data sets in the future.

A preliminary version of this work was published in Nalluri et. al, 2013[53] and thereafter a secondary version in Nalluri et. al, 2017[125]

CHAPTER 7

MIRNA-SNP INTERACTIONS IN DISEASES

7.1 Motivation

The pathogenesis of many diseases are often linked to genetic factors. Understanding these genetic factors are crucial for diagnosis and treatment of these diseases; especially those which are multifactoral in nature and arising out of multiple culpable interactions. Single-nucleotide polymorphism (SNP) is defined as a variation in a single nucleotide that occurs at a specific location(position) in a DNA sequence/genome. They are the most common form of DNA sequence variation, and are responsible for genetic differences between individuals. These SNPs occur in genes, non-coding regions of genes or in regions between the genes. SNPs occuring in the coding regions are classified as synonymous and non-synonymous. Non-synonymous SNPs are known to cause a change in the amino acid sequences which result in the alteration of the structure or function of a protein. Genome-wide studies (GWAS) and expression quantitative trait loci (eQTLs) studies have identified several SNPs associated with complex diseases.

miRNAs are known to be key regulators of gene expression via miRNA-mRNA binding. SNPs in the miRNA gene or in the binding site of a target mRNA can derail the gene's regulatory mechanism by either destroying or creating a new binding site. If a new miRNA target site is created, it decreases the mRNA translation and if an existing miRNA target site is destroyed, there is an increase in mRNA translation. Hence, exploring the role of SNPs in a miRNA regulatory network is crucial to understand its consequential effects in diseases.

7.2 Materials and Methods

In this work, we aim to study the impact SNPs have in the miRNA's regulatory network, pertaining to diseases. In Chapter 5, we created disease-specific miRNA-miRNA interaction networks (DMINs) pertaining to several diseases. In Chapter 6, we determined the causal miRNAs among four categories of tissue-specific cancers. In this work, we augment the existing DMINs, with the regulatory role initiated by SNPs and SNP-related miRNAs.

An overview of the methodology is shown in Figure 35.

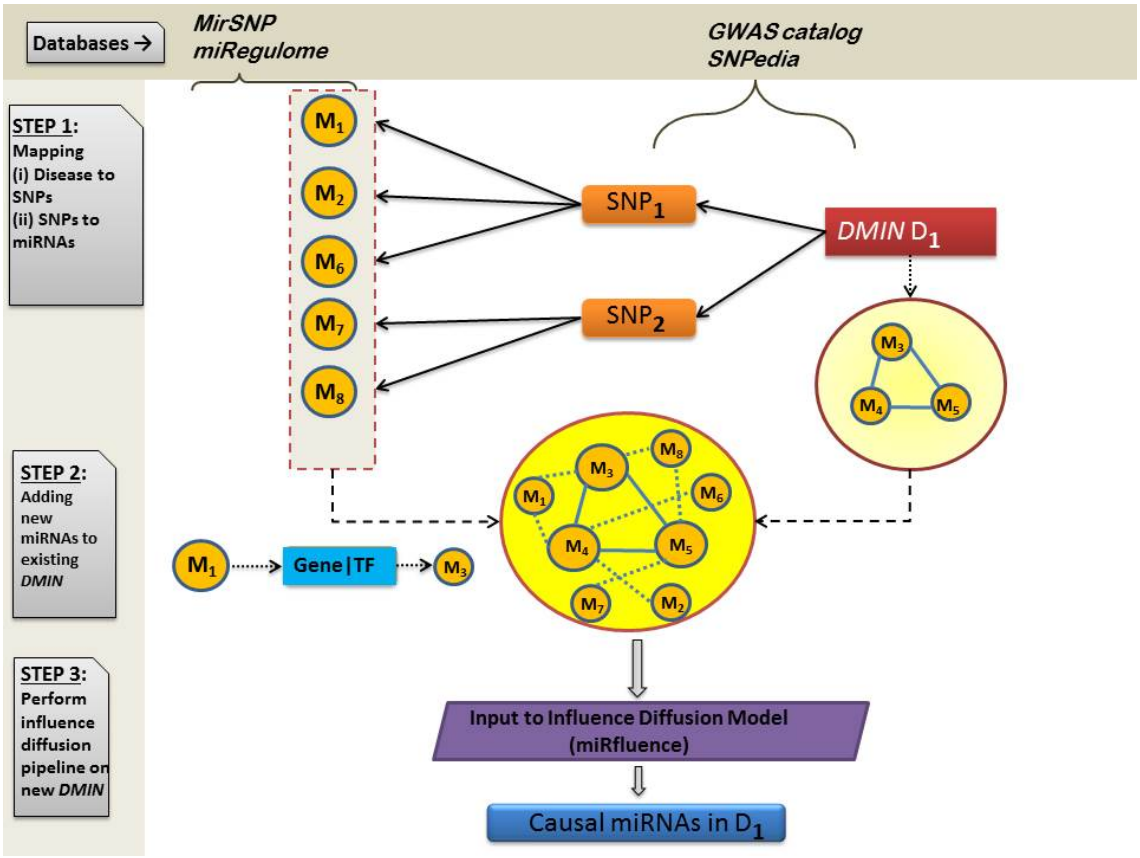


Fig. 35.: Overview of the miRNA-SNP methodology

1. Step 1:

(A) : To map the associations between SNPs and their associated diseases. Several databases like *SNPedia*[126], GWAS catalog[127] have documented these associations. These associations are manually curated and have high degree of confidence scores. *SNPedia* provides a structured and curated form of available literature citing the SNP based associations. It summarizes the medical, phenotypic and genealogical associations of these DNA variations. Many DNA testing services, such as *23andMe*[128], *Ancestry*[129], *FamilyTreeDNA*[130] use *SNPedia* based services to learn more about their DNA variants. GWAS catalog is a collection of more than 100,000 SNPs and their trait associations with high confidence scores and p-values of likelihood, published by the National Human Genome Research Institute-EBI. The GWAS catalog contains data on about 11,912 SNPs curated from 1751 publications[127]. Using these resources, we map/determine the SNPs associated to risks in diseases, via genes. (see Figure 35, Step 1)

(B): Here, we map the associations between the discovered SNPs (from the previous step) and miRNAs. We use *miRSNP*[131] to map these associations. *miRSNP* is a database of SNPs altering miRNA-target sites and miRNA genes. *miRSNP* has 414,510 SNPs that affect miRNA-mRNA binding. These associations are further annotated with information whether a SNP in the mRNA target site would decrease/break or increase/create affinity. It identifies miRNA-related SNPs. We use this dataset to determine the SNP-related miRNAs. The data was curated from *miRBase*[8] and *dbSNP*[132] for miRNAs and SNPs respectively.

(C) : Here, we map the associations between TFs and miRNAs. We use databases *miRegulome*[51] (Chapter 3), *miRSNP*[131] to map these associa-

tions. *miRSNP* is a database of SNPs altering miRNA-target sites and miRNA genes. It also identifies miRNA-related SNPs. This information supplements the data from *SNP2TBFS*s. We use these datasets to determine the SNP-related miRNAs and TF-regulated miRNAs.

2. Step 2: The miRNAs identified by Steps 1(A-B) are appended to the existing **d**isease-specific **m**iRNA **i**nteraction **n**etwork (DMIN), Figure 35, Step 2). These new miRNAs are embedded with the information from SNP based interactions pertaining to the diseases, and hence are crucial in supplementing the existing miRNAs in *DMIN*s. The challenge however, is in determining the edge interactions among the newly discovered SNP-related miRNAs and existing disease-specific miRNAs in the *DMIN*.

   As miRNAs regulate target genes, we determine the target genes which are specifically regulated by SNP-related miRNAs. Many of the target genes also act as transcription factors (TFs). Therefore, we identify the target genes under the regulatory control of SNP-related miRNAs which are also acting as TFs to the existing disease-specific miRNAs in the *DMIN*. This is demonstrated in Figure 35, Step II. Hence, if there is an interaction found, such that a SNP-related miRNA is regulating a target gene, which in turn is acting as a TF and regulating a disease-specific miRNA, we consider this interaction as an edge between the SNP-related miRNA and the existing disease-specific miRNA.

   We perform this querying among all the SNP-related miRNAs and disease-specific miRNAs and add the new list of interactions (edges) into the existing *DMIN*, thereby, augmenting it with SNP-related information.

3. Step 3: After Step 2, the newly created *DMIN*s contain information about miRNA-miRNA interactions emerging from both — fold change values and

SNP-initiated ramifications, and hence provide a richer sense of the data. We implement the influence diffusion model (created in Chapter 6) to determine the causal miRNA nodes in the newly created *DMIN* for every disease and across other disease categories by combining the *DMINs* of similar diseases.

7.3 Results

We implemented this methodology on two disease datasets - chronic lymphatic leukemia (CLL) and hematological tumors. In order to compare the efficacy of the proposed methodology, we deployed the influence diffusion model on the *DMIN*, before and after adding the the SNP-related miRNA-miRNA edges.

1. Chronic Lymphatic Leukemia (CLL)

   In Table 10, the left-side table represents the causal miRNAs in the original existing DMIN and the right-side table represents the causal miRNAs in the *DMIN* integrated with SNP-related edges. As demonstrated in Table 10, after adding the SNP-based interactions into the existing *DMIN*, a SNP-related miRNA was implicated to be causal for CLL.

2. Hematological tumors

   In Table 11, the left-side table represents the causal miRNAs in the original existing DMIN and the right-side table represents the causal miRNAs in the *DMIN* integrated with SNP-related edges. As demonstrated in Table 10, after adding the SNP-based interactions into the existing *DMIN*, a SNP-related miRNA was implicated to be causal for *hematological tumors*.

In this work, as demonstrated in the *Results*, we incorporate the SNP-based interactions and information into the existing disease-specific miRNA-miRNA interaction networks and are able to identify SNP-related miRNAs which play a causal

role in the pathogenesis of chronic lymphatic leukemia and hematological tumors. The novel contribution lies in the construction of the *DMIN* with SNP-information embedded into the miRNAs and also the introduction of SNP-based miRNA-miRNA interactions. SNPs have most often led to causal roles in the genesis of diseases. Using our influence diffusion model in these enhanced *DMIN*s, we are able to identify causal miRNAs for the diseases of chronic lymphatic leukemia and hematological tumors. It should be noted that, due to the unavailability of common diseases between the GWAS catalog and our datasets, the results were limited to the aforementioned diseases.

Table 10.: *Left*: Top 10 causal miRNAs in *DMIN* of Chronic Lymphatic Leukemia, **without** the SNP-related miRNAs added. *Right:* Causal miRNAs in *DMIN* of Chronic Lymphatic Leukemia, **with** the SNP-related miRNAs added. The results showed an implication of a SNP-related miRNA to be causal (highlighted in blue)

Rank	Causal miRNAs
1	hsa-mir-181c
2	hsa-mir-130b
3	hsa-mir-130a
4	hsa-mir-128-2
5	hsa-mir-125b-1
6	hsa-mir-92a-1
7	hsa-mir-126
8	hsa-mir-100
9	hsa-mir-7b
10	hsa-mir-99a

Rank	Causal miRNAs
1	hsa-mir-181c
2	hsa-mir-130b
3	hsa-mir-126
4	hsa-mir-130a
5	hsa-mir-27a
6	hsa-mir-125b-1
7	hsa-mir-92a-1
8	hsa-mir-128-2
9	hsa-mir-100
10	hsa-mir-7b

Table 11.: *Left*: Top 10 causal miRNAs in *DMIN* of hematological tumors, **without** the SNP-related miRNAs added. *Right:* Causal miRNAs in *DMIN* of hematological tumors, **with** the SNP-related miRNAs added. The results showed an implication of two SNP-related miRNAs to be causal (highlighted in blue) among the top 10.

Rank	Causal miRNAs
1	hsa-mir-219-2
2	hsa-mir-200a
3	hsa-mir-23a
4	hsa-mir-379
5	hsa-mir-212
6	hsa-mir-338
7	hsa-mir-125b-1
8	hsa-mir-370
9	hsa-mir-190
10	hsa-mir-24-1

Rank	Causal miRNAs
1	hsa-mir-23a
2	hsa-mir-200a
3	hsa-mir-379
4	hsa-mir-149
5	hsa-mir-219-2
6	hsa-mir-133b
7	hsa-mir-92a-1
8	hsa-mir-125b-1
9	hsa-mir-30b
10	hsa-mir-212

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

In this dissertation, we reported several computational methodologies conceived and implemented incrementally and applied to determine the patterns of miRNA regulomic behavior and also show-case the interactions in miRNA-disease networks. We study these interaction patterns and regulatory behavior especially in the context of diseases. Each of the reported methodologies has been made available to the broader research community via online softwares and web-tools. Our future work involves integrating newer components of the miRNA regulome as they become available into our integrated framework; equally important is the design of integrated analytical tools that can enable systems level hypothesis making. Moreover, our integrated miRNA regulomics platform can be used to drive future research in several new directions as follows:

- Synthetic biology: Due to the efficient communication properties of feed-forward loops (FFLs), specifically in terms of noise filtering and robust signal transport, they make great candidates for the emerging domain of synthetic biology where larger engineered TRN circuits can be built that are resilient to external perturbations[133, 134]. Early efforts in this direction have shown great promise and the importance of integrated miRNA-TF-gene regulatory networks (that form individual FFLs) as investigated here can motivate the construction of more efficient genetic circuits in the future.

- Biological network growing algorithms: Another popular area of research includes the transcriptional network growing algorithms primarily based on the

preferential attachment model [135] or its variations [136]. Currently, only the TRNs of Escherichia coli (*E. coli*) and Saccharomyces cerevisiae (*Yeast*) have been validated experimentally; hence such network growing algorithms are essential to allow the community to study the properties of such TRNs for designing robust networks [137], as well as to predict the TRNs of higher-level organisms. Future studies on the topologies of the integrated miRNA-TF-gene regulatory networks can provide new directions in this line of research.

- Bio-inspired Wireless Sensor Networking: Wireless sensor networks form a special class of engineered systems wherein sensor nodes forward data packets that are routed through adjacent sensors to a sink capable of processing the sensed information. Resemblance between gene regulation systems and wireless sensor networks (herein WSNs) can be described through transcription, where genes process signals from adjacent neighbors in the form of transcription factors that excite/repress other genes by generating mRNA molecules. Nodes in a TRN interface by conveying signals (transcription factors), that are then processed into output signals (mRNAs). WSNs operate in a similar manner, where sensor nodes send signals to others in the form of data packets. Packets at destination nodes convey forwarding instructions, which in return relays such packets to other sensors. Recent studies show that wireless sensor networks adopting the transcriptional regulatory topologies (of E. coli), designated as bio-inspired WSNs, are more efficient than those adopting random network topologies of the same size in terms of conveying packets to sink nodes [137–150]. This thesis will motivate the design of new smart WSN topologies by exploiting the integrated regulatory networks reported here that may exhibit better efficiency in terms of their average packet receipt rates under node/link failures and channel noise.

Appendix A

DETAILS OF NETWORK INFERENCE ALGORITHMS

These network inference algorithms were used in Chapter 5.

1. CLR

   Context likelihood of relatedness (CLR)[71] algorithm belongs to the class of relevance networks. Relevance network algorithms use mutual information (MI) based Z scores between a regulator-target pair to identify and determine potential interactions between them. If the MI score is above a certain threshold among the expression dataset, the interaction between the pair is highly likely. In the past, CLR method has been proven to be effective in learning novel transcriptional interactions in E. coli[71]. CLR uses the metric of MI to gauge the similarity between the expression profiles of two entities, in our case, disease-specific miRNAs. The MI is calculated as below:

$$I(X,Y) = \sum_{i,j} P(x_i, y_i) log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \tag{A.1}$$

   where, X and Y are random variables i.e. two miRNAs in this case. $P(x_i)$ denotes the probability of X = x_i. MI values between a $miRNA_i$ and $miRNA_j$ are calculated and thereafter estimated with regards to the likelihood of their occurrence by comparing that score with a background null model, which is the distribution of MI values. The null model incorporates two sets of MI values: MI_i which is a set of all $miRNA_i$'s MI values and MI_j, a set of all MI values of $miRNA_j$. These two sets, MI_i and MI_j are two independent variables used in

the computation of the joint distribution of the null-model, i.e. the background MI. Thus the MI score of the pair $(miRNA_i, miRNA_j)$ is compared with two Z scores resulting from MI_i and MI_j. Further in-depth explanation of the algorithm can be found in[71]. However, CLR predominantly relies on the MI matrix for its scores, and cannot ascertain for causality between the regulators which is more often based on the regulatory kinetics exhibited in the time-series data. The CLR algorithm, available in the 'minet' R Bioconductor package was used in this work.

2. GENIE3

   Gene Network Inference with Ensemble of Trees (GENIE3)[74] algorithm was the top performing algorithm in the 'DREAM4- In Silico Network Challenge' of inferring gene regulatory network. GENIE3 has a different approach of inferring the associations by taking into account, feature selection as compared to CLR, which is mutual-information based. In GENIE3, a regression problem is formulated for each individual miRNA. Hence in our case, 4343 regression problems were solved for each disease-specific miRNAs. In this regression analysis, each disease-specific miRNA's expression pattern is predicted from every other miRNA's expression pattern by the application of 'Random Forests' model which is a tree-based ensemble model. Thus, based on the significance of the expression patterns between two disease-specific miRNAs and the least output variance between the target miRNA and the considered miRNA, a regulatory link between them is predicted. In this way, the algorithm ranks all the interactions between the miRNAs based on their aggregated scores from regression analysis. GENIE3 is adaptable to other categories of expression data involving interactions. The GENIE algorithm hosted on GenePattern website was exe-

cuted with standard run conditions - tree-based method as Random Forests' and the number of trees grown in an ensemble as 500.

3. Basic Correlation

   The Basic Correlation method ranks every disease-specific miRNA-miRNA pair according to the correlation between them. This algorithm uses the Pearsons and Spearmans coefficient to calculate their correlation score,

   Pearsons coefficient:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\rho_X \rho_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\rho_X \rho_Y} \tag{A.2}$$

   where, X and Y are random variables, with μ_X and μ_Y being the expected values and ρ_X and ρ_Y being the standard deviations. The Spearman correlation coefficient

$$\rho = 1 - \frac{6\sum(d_i)^2}{n(n^2-1)} \tag{A.3}$$

   where, d_i is the difference between the ranks of corresponding values X_i and Y_i and n is the number of points in dataset. While running the Basic Correlation algorithm hosted on the tool GenePattern the ranks of the disease-specific miRNA-miRNA were derived using Pearson's correlation coefficient and Spearman's correlation. Note that, there are two resultant files - one with Pearson's correlation coefficient and the other with Spearman's correlation coefficient, under this approach.

4. MRNETB

   MRNETB[72] is a mutual information based network inference algorithm which

is an improved version of its predecessor MRNET. MRNET performs network inference using the method 'Maximum Relevance Minimum Redundancy (MRMR)'. For every random variable X_i, a set of predictor variables are chosen based on the difference of mutual information between X_i and the set of variables $X_i \in X_{S_j}$. This method follows forward selection, i.e. a variable X_i with highest MI score with the target variable X_{S_j} is chosen at first. Hence, the general idea behind the algorithm is identification of subsets of X_{S_j} for every variable with which the variable in the set has maximum pairwise relevance and maximum pairwise independence. While previous methods recursively select a subset of variables and compute the mutual information scores, MRNETB performs backward elimination with sequential search. Further, in-depth explanation can be found at[72].

This algorithm was run from the Bioconductor package minet with Spearman entropy estimator and the number of bins used for discretization was $\sqrt{N}$, where N=number of samples, i.e. 267 in our expression dataset.

5. Distance Correlation

   This algorithm is described in[73] which uses a novel measurement of dependence called distance correlation (DC) to derive non-linear dependencies from the gene expression dataset. This metric uses a different approach as compared to some of the previous MI based methods. MI based methods rely on density estimator of certain patterns which can be challenging for multivariate data, it cab be challenging. Also, in the case of continuous data, it has to be discretized before MI based methods are applied. The algorithm details can be found in their work[73] and are beyond the scope of this work.

PUBLICATIONS

Journals

1. Nalluri Joseph J., Pratip Rana, Debmalya Barh, Vasco Azevedo, Thang N. Dinh, Vladimir Vladimirov and Preetam Ghosh. "Determining causal miRNAs and their signaling cascade in diseases using an influence diffusion model." *Scientific reports* 7 (2017)**(IF: 4.25)**

2. Nalluri, Joseph J., Debmalya Barh, Vasco Azevedo, and Preetam Ghosh. "miRsig: a consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures." *Scientific reports* 7 (2017).**(IF: 4.25)**

3. Nalluri, Joseph J., D. Barh, V. Azevedo, and P. Ghosh. "Towards a Comprehensive Understanding of Mirna Regulome and Mirna Interaction Networks." J Pharmacogenomics Pharmacoproteomics 7, no. 160 (2016): 2153-0645.

4. Barh, Debmalya, Bhanu Kamapantula*, Neha Jain, Joseph Nalluri*, Antaripa Bhattacharya, Lucky Juneja, Neha Barve et al. "miRegulome: a knowledgebase of miRNA regulomics and analysis." *Scientific reports* 5 (2015): 12832.**(IF: 4.25)** **Equal contributor*

5. Nalluri, Joseph J., Bhanu K. Kamapantula, Debmalya Barh, Neha Jain, Antaripa Bhattacharya, Sintia Silva de Almeida, Rommel Thiago Juca Ramos, Artur Silva, Vasco Azevedo, and Preetam Ghosh. "DISMIRA: Prioritization of disease candidates in miRNA-disease associations based on maximum weighted matching inference model and motif-based analysis." *BMC Genomics* 16, no. 5 (2015): S12.**(IF: 3.86)**

Conference Proceedings

1. Nalluri, Joseph, Pratip Rana, Vasco Azevedo, Debmalya Barh, and Preetam Ghosh. "Determining influential miRNA targets in diseases using influence diffusion model." In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 519-520. ACM, 2015.

2. Nalluri, Joseph, Bhanu Kamapantula, Preetam Ghosh, Debmalya Barh, Neha Jain, Lucky Juneja, and Neha Barve. "Determining mirna-disease associations using bipartite graph modelling." In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, p. 672. ACM, 2013.

Book Chapters

1. Nalluri Joseph J., Debmalya Barh, Vasco Azevedo and Preetam Ghosh, "Bioinformatics and systems biology in bio-engineering," *Omics Technologies and Bioengineering: Towards Improving Quality of Life*. Ed. D. Barh. Academic Press, 2017. Print.

2. Debmalya Barh, Eugenia Ch Yiannakopoulou, Emmanuel O Salawu, Atanu Bhattacharjee, Sudhir Chowbina, Joseph J. Nalluri, Preetam Ghosh, Vasco Azevedo, "In Silico Disease Model: From Simple Networks to Complex Diseases" Animal Biotechnology: Models in discovery

REFERENCES

[1] Witold Filipowicz, Suvendra N Bhattacharyya, and Nahum Sonenberg. "Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?" In: *Nature Reviews Genetics* 9.2 (2008), pp. 102–114.

[2] Tsuyoshi Yokoi and Miki Nakajima. "Toxicological implications of modulation of gene expression by microRNAs". In: *Toxicological Sciences* 123.1 (2011), pp. 1–14.

[3] Azra Krek et al. "Combinatorial microRNA target predictions". In: *Nature genetics* 37.5 (2005), pp. 495–500.

[4] Qinghua Jiang et al. "Predicting human microRNA-disease associations based on support vector machine". In: *Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on*. IEEE. 2010, pp. 467–472.

[5] Ping Xuan et al. "Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors". In: *PloS one* 8.8 (2013), e70204.

[6] Qinghua Jiang et al. "Weighted Network-Based Inference of Human MicroRNA-Disease Associations". In: *Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on*. IEEE. 2010, pp. 431–435.

[7] Hailin Chen and Zuping Zhang. "Similarity-based methods for potential human microRNA-disease association prediction". In: *BMC medical genomics* 6.1 (2013), p. 12.

[8] Sam Griffiths-Jones et al. "miRBase: microRNA sequences, targets and gene nomenclature". In: *Nucleic acids research* 34.suppl 1 (2006), pp. D140–D144.

[9] Qinghua Jiang et al. “miR2Disease: a manually curated database for microRNA deregulation in human disease”. In: *Nucleic acids research* 37.suppl 1 (2009), pp. D98–D104.

[10] Juan Wang et al. “TransmiR: a transcription factor–microRNA regulation database”. In: *Nucleic acids research* 38.suppl 1 (2010), pp. D119–D122.

[11] Qingqing Yang et al. “miREnvironment database: providing a bridge for microRNAs, environmental factors and phenotypes”. In: *Bioinformatics* 27.23 (2011), pp. 3329–3330.

[12] Feifei Xiao et al. “miRecords: an integrated resource for microRNA–target interactions”. In: *Nucleic acids research* 37.suppl 1 (2009), pp. D105–D110.

[13] Harsh Dweep et al. “miRWalk database: prediction of possible miRNA binding sites by walking the genes of three genomes”. In: *Journal of biomedical informatics* 44.5 (2011), pp. 839–847.

[14] Elize A Shirdel et al. “NAViGaTing the micronome–using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs”. In: *PloS one* 6.2 (2011), e17429.

[15] Sheng-Da Hsu et al. “miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions”. In: *Nucleic acids research* 42.D1 (2014), pp. D78–D85.

[16] Natalia J Martinez et al. “A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity”. In: *Genes & development* 22.18 (2008), pp. 2535–2549.

[17] Matteo Osella et al. “The role of incoherent microRNA-mediated feedforward loops in noise buffering”. In: *PLoS Comput Biol* 7.3 (2011), e1001101.

[18] Qinghua Jiang, Guohua Wang, and Yadong Wang. "An approach for prioritizing disease-related microRNAs based on genomic data integration". In: *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*. Vol. 6. IEEE. 2010, pp. 2270–2274.

[19] Xing Chen, Ming-Xi Liu, and Gui-Ying Yan. "RWRMDA: predicting novel human microRNA–disease associations". In: *Molecular BioSystems* 8.10 (2012), pp. 2792–2798.

[20] Ming Lu et al. "An analysis of human microRNA and disease associations". In: *PloS one* 3.10 (2008), e3420.

[21] Juan Xu et al. "Prioritizing candidate disease miRNAs by topological features in the mirna target–dysregulated network: Case study of prostate cancer". In: *Molecular cancer therapeutics* 10.10 (2011), pp. 1857–1866.

[22] Simona Rossi et al. "OMiR: Identification of associations between OMIM diseases and microRNAs". In: *Genomics* 97.2 (2011), pp. 71–76.

[23] Zuping Zhang. "Prediction of Associations between OMIM Diseases and MicroRNAs by Random Walk on OMIM Disease Similarity Network". In: *The Scientific World Journal* 2013 (2013).

[24] Xing Chen and Gui-Ying Yan. "Semi-supervised learning for potential human microRNA-disease associations inference". In: *Scientific reports* 4 (2014).

[25] Dong Wang et al. "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases". In: *Bioinformatics* 26.13 (2010), pp. 1644–1650.

[26] Jie Li et al. "Computational prediction of microRNA networks incorporating environmental toxicity and disease etiology". In: *Scientific reports* 4 (2014).

[27] Feixiong Cheng et al. "Prediction of drug-target interactions and drug repositioning via network-based inference". In: *PLoS computational biology* 8.5 (2012), e1002503.

[28] Dandan Yuan et al. "Enrichment Analysis Identifies Functional MicroRNA-Disease Associations in Humans". In: *PloS one* 10.8 (2015), e0136285.

[29] Quan Zou et al. "Prediction of microRNA-disease associations based on social network analysis methods". In: *BioMed research international* 2015 (2015), p. 810514.

[30] Bo Liao et al. "Identifying human microRNA–disease associations by a new diffusion-based method". In: *Journal of bioinformatics and computational biology* 13.04 (2015), p. 1550014.

[31] Xing Chen et al. "WBSMDA: Within and Between Score for MiRNA-Disease Association prediction". In: *Scientific reports* 6 (2016).

[32] Hongbo Shi et al. "Integration of Multiple Genomic and Phenotype Data to Infer Novel miRNA-Disease Associations". In: *PloS one* 11.2 (2016), e0148521.

[33] Xing Chen et al. "RBMMMDA: predicting multiple types of disease-microRNA associations". In: *Scientific reports* 5 (2015).

[34] Juan Xu et al. "MiRNA–miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features". In: *Nucleic acids research* 39.3 (2011), pp. 825–836.

[35] Renhua Song et al. "Identification of lung cancer miRNA-miRNA co-regulation networks through a progressive data refining approach". In: *Journal of Theoretical Biology* (2015).

[36] Chaohan Xu et al. "Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles". In: *Mol. BioSyst.* 10.11 (2014), pp. 2800–2809.

[37] Chen-Ching Lin et al. "A cross-cancer differential co-expression network reveals microRNA-regulated oncogenic functional modules". In: *Molecular BioSystems* 11.12 (2015), pp. 3244–3252.

[38] Sanghamitra Bandyopadhyay et al. "Development of the human cancer microRNA network". In: *Silence* 1.1 (2010), p. 1.

[39] Joseph J Nalluri et al. "DISMIRA: Prioritization of disease candidates in miRNA-disease associations based on maximum weighted matching inference model and motif-based analysis". In: *BMC genomics* 16.Suppl 5 (2015), S12.

[40] Sungroh Yoon and Giovanni De Micheli. "Prediction of regulatory modules comprising microRNAs and target genes". In: *Bioinformatics* 21.suppl 2 (2005), pp. ii93–ii100.

[41] Georgios A Pavlopoulos et al. "Using graph theory to analyze biological networks". In: *BioData mining* 4.1 (2011), p. 1.

[42] Hyunghoon Cho, Bonnie Berger, and Jian Peng. "Reconstructing Causal Biological Networks through Active Learning". In: *PloS one* 11.3 (2016), e0150611.

[43] Niranjan Nagarajan and Carl Kingsford. "GiRaF: robust, computational identification of influenza reassortments via graph mining". In: *Nucleic acids research* (2010), gkq1232.

[44] Amrita Pati et al. "XcisClique: analysis of regulatory bicliques". In: *BMC bioinformatics* 7.1 (2006), p. 1.

[45] Eytan Bakshy et al. “The role of social networks in information diffusion”. In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 519–528.

[46] Zsolt Katona, Peter Pal Zubcsek, and Miklos Sarvary. “Network effects and personal influences: The diffusion of an online social network”. In: *Journal of Marketing Research* 48.3 (2011), pp. 425–443.

[47] Dunia López-Pintado. “Diffusion in complex social networks”. In: *Games and Economic Behavior* 62.2 (2008), pp. 573–590.

[48] Jacqueline Johnson Brown and Peter H Reingen. “Social ties and word-of-mouth referral behavior”. In: *Journal of Consumer research* (1987), pp. 350–362.

[49] Mark Granovetter. “Threshold models of collective behavior”. In: *American journal of sociology* (1978), pp. 1420–1443.

[50] Jing Chen et al. “ToppGene Suite for gene list enrichment analysis and candidate gene prioritization”. In: *Nucleic acids research* 37.suppl 2 (2009), W305–W311.

[51] Debmalya Barh et al. “miRegulome: a knowledge-base of miRNA regulomics and analysis”. In: *Scientific reports* 5 (2015).

[52] Joseph J. Nalluri et al. “Towards a Comprehensive Understanding of miRNA Regulome and miRNA Interaction Networks”. In: *Journal of Pharmacogenomics and Pharmacoproteomics* 7.160 (2016), pp. 2153–0645.

[53] Joseph Nalluri et al. “Determining miRNA-disease associations using bipartite graph modelling”. In: *Proceedings of the International Conference on Bioin-*

formatics, Computational Biology and Biomedical Informatics. ACM. 2013, p. 672.

[54] Armen S Asratian. *Bipartite graphs and their applications*. 131. Cambridge University Press, 1998.

[55] Andrew Makhorin. "{GNU} linear programming kit". In: *Moscow Aviation Institute, Moscow, Russia* 38 (2001).

[56] Ron Milo et al. "Network motifs: simple building blocks of complex networks". In: *Science* 298.5594 (2002), pp. 824–827.

[57] Nadav Kashtan et al. "Mfinder tool guide". In: *Department of Molecular Cell Biology and Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot Israel, Tech. Rep* (2002).

[58] Sebastian Wernicke and Florian Rasche. "FANMOD: a tool for fast network motif detection". In: *Bioinformatics* 22.9 (2006), pp. 1152–1153.

[59] J-P Onnela et al. "Structure and tie strengths in mobile communication networks". In: *Proceedings of the National Academy of Sciences* 104.18 (2007), pp. 7332–7336.

[60] Documentación Django. "en l{\'\i}nea]: Django The Web Framework for perfectionists with deadlines". In: *Culver City. California. Estados Unidos* (2007).

[61] Michel F Sanner et al. "Python: a programming language for software integration and development". In: *J Mol Graph Model* 17.1 (1999), pp. 57–61.

[62] A Hagberg, D Schult, and P Swart. "NetworkX". In: *URL http://networkx. github. io/index. html* (2013).

[63] David Flanagan. "JavaScript: the definitive guide". In: O'Reilly Media, Inc., 2002.

[64] Michael Bostock. "D3. js". In: *Data Driven Documents* (2012).

[65] Rui Tang et al. "Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system". In: *Cell research* 22.3 (2012), pp. 504–515.

[66] Bing Shi et al. "Highly ordered architecture of microRNA cluster". In: *BioMed research international* 2013 (2013).

[67] Mark P Hamilton et al. "Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif". In: *Nature communications* 4 (2013).

[68] Andreas Ruepp et al. "PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes". In: *Genome biology* 11.1 (2010), R6.

[69] James Surowiecki. *The wisdom of crowds*. Anchor, 2005.

[70] Daniel Marbach et al. "Wisdom of crowds for robust gene network inference". In: *Nature methods* 9.8 (2012), pp. 796–804.

[71] Jeremiah J Faith et al. "Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles". In: *PLoS biology* 5.1 (2007), e8.

[72] Patrick Meyer et al. "Information-Theoretic Inference of Gene Networks Using Backward Elimination." In: *BIOCOMP*. 2010, pp. 700–705.

[73] Xiaobo Guo et al. "Inferring Nonlinear Gene Regulatory Networks from Gene Expression Data Based on Distance Correlation". In: *PloS one* 9.2 (2014), e87446.

[74] Alexandre Irrthum, Louis Wehenkel, Pierre Geurts, et al. "Inferring regulatory networks from expression data using tree-based methods". In: *PloS one* 5.9 (2010), e12776.

[75] Michael Bostock. *Data-Driven Documents*. `http:https://d3js.org/`. Accessed: 2016-04-18.

[76] Fabian Afonso-Grunz and Sören Müller. "Principles of miRNA–mRNA interactions: Beyond sequence complementarity". In: *Cellular and Molecular Life Sciences* 72.16 (2015), pp. 3127–3141.

[77] Li Guo et al. "miRNA–miRNA interaction implicates for potential mutual regulatory pattern". In: *Gene* 511.2 (2012), pp. 187–194.

[78] S Arora et al. "miRNA–transcription factor interactions: a combinatorial regulation of gene expression". In: *Molecular genetics and genomics* 288.3-4 (2013), pp. 77–87.

[79] Ali M Ardekani and Mozhgan Moslemi Naeini. "The role of microRNAs in human diseases". In: *Avicenna journal of medical biotechnology* 2.4 (2011), pp. 161–180.

[80] Lin Hua et al. "Combination of microRNA expression profiling with genome-wide SNP genotyping to construct a coronary artery disease-related miRNA-miRNA synergistic network". In: *Bioscience trends* 8.6 (2014), pp. 297–307.

[81] Bingli Wu et al. "Dissection of miRNA-miRNA interaction in esophageal squamous cell carcinoma". In: *PloS one* 8.9 (2013), e73191.

[82] Mohammed Alshalalfa. "MicroRNA response elements-mediated miRNA-miRNA interactions in prostate cancer". In: *Advances in bioinformatics* 2012 (2012).

[83] Mariano Monzo et al. "Overlapping expression of microRNAs in human embryonic colon and colorectal cancer". In: *Cell research* 18.8 (2008), pp. 823–833.

[84] Tetsuya Ueda et al. "Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis". In: *The lancet oncology* 11.2 (2010), pp. 136–146.

[85] Yuqing Zhang et al. "Profiling of 95 microRNAs in pancreatic cancer cell lines and surgical specimens by real-time PCR analysis". In: *World journal of surgery* 33.4 (2009), pp. 698–709.

[86] Hushan Yang et al. "MicroRNA expression signatures in Barrett's esophagus and esophageal adenocarcinoma". In: *Clinical Cancer Research* 15.18 (2009), pp. 5744–5752.

[87] Yong Guo et al. "Distinctive microRNA profiles relating to patient survival in esophageal squamous cell carcinoma". In: *Cancer research* 68.1 (2008), pp. 26–33.

[88] Sheng-Li Zhou and Li-Dong Wang. "Circulating microRNAs: novel biomarkers for esophageal cancer". In: *World J Gastroenterol* 16.19 (2010), pp. 2348–2354.

[89] Jinmai Jiang et al. "Association of MicroRNA expression in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival". In: *Clinical Cancer Research* 14.2 (2008), pp. 419–427.

[90] Claudia Roldo et al. "MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic fea-

tures and clinical behavior". In: *Journal of Clinical Oncology* 24.29 (2006), pp. 4677–4684.

[91] Marina N Nikiforova et al. "MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility". In: *The Journal of Clinical Endocrinology & Metabolism* 93.5 (2008), pp. 1600–1608.

[92] Jinmai Jiang et al. "Real-time expression profiling of microRNA precursors in human cancer cell lines". In: *Nucleic acids research* 33.17 (2005), pp. 5394–5403.

[93] Shadan Ali et al. "Differentially expressed miRNAs in the plasma may provide a molecular signature for aggressive pancreatic cancer". In: *Am J Transl Res* 3.1 (2010), pp. 28–47.

[94] Ming-Zhe Ma et al. "Candidate microRNA biomarkers of pancreatic ductal adenocarcinoma: meta-analysis, experimental validation and clinical significance". In: *Journal of Experimental & Clinical Cancer Research* 32.1 (2013), p. 1.

[95] Petra Leidinger, Andreas Keller, and Eckart Meese. "MicroRNAs–important molecules in lung cancer research". In: *Frontiers in genetics* 2 (2012), p. 104.

[96] D L Zanette et al. "miRNA expression profiles in chronic lymphocytic and acute lymphocytic leukemia". In: *Brazilian Journal of Medical and Biological Research* 40.11 (2007), pp. 1435–1440.

[97] Ramiro Garzon et al. "MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia". In: *Blood* 111.6 (2008), pp. 3183–3189.

[98] Ramiro Garzon et al. "Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin". In: *Proceedings of the National Academy of Sciences* 105.10 (2008), pp. 3945–3950.

[99] James Andrew Sutherland Colquhoun. *With the Kurram Field Force, 1878-79*. WH Allen & Company, 1881.

[100] Jorge R Contreras et al. "MicroRNA-146a modulates B-cell oncogenesis by regulating Egr1". In: *Oncotarget* 6.13 (2015), p. 11023.

[101] Amanda J Favreau et al. "miR-199b, a novel tumor suppressor miRNA in acute myeloid leukemia with prognostic implications". In: *Experimental hematology & oncology* 5.1 (2016), p. 1.

[102] Robert E Donahue et al. "Plerixafor (AMD3100) and granulocyte colony-stimulating factor (G-CSF) mobilize different CD34+ cell populations based on global gene and microRNA expression signatures". In: *Blood* 114.12 (2009), pp. 2530–2541.

[103] Bastiaan JH Jansen et al. "MicroRNA genes preferentially expressed in dendritic cells contain sites for conserved transcription factor binding motifs in their promoters". In: *BMC genomics* 12.1 (2011), p. 1.

[104] Sanjay K Singh et al. "A microRNA link to glioblastoma heterogeneity". In: *Cancers* 4.3 (2012), pp. 846–872.

[105] Xiwei Wu et al. "Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis". In: *PloS one* 7.5 (2012), e35661.

[106] Sanghamitra Bandyopadhyay and Malay Bhattacharyya. "Analyzing miRNA co-expression networks to explore TF-miRNA regulation". In: *BMC bioinformatics* 10.1 (2009), p. 163.

[107] Zhenjun Hu et al. "VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies". In: *Nucleic acids research* 41.W1 (2013), W225–W231.

[108] Xia Lab. *miRNet, network-based visual analysis of miRNAs, targets and functions*. \url{http://www.mirnet.ca/}. 2015.

[109] Joseph Nalluri et al. "Determining causal miRNAs and their signaling cascade in diseases using an influence diffusion model". In: *Scientific reports* 7 (2017).

[110] Lada A Adamic and Eytan Adar. "Friends and neighbors on the web". In: *Social networks* 25.3 (2003), pp. 211–230.

[111] Zoltán Dezső and Albert-László Barabási. "Halting viruses in scale-free networks". In: *Physical Review E* 65.5 (2002), p. 055103.

[112] Mark DF Shirley and Steve P Rushton. "The impacts of network topology on disease spread". In: *Ecological Complexity* 2.3 (2005), pp. 287–299.

[113] Wei Wang et al. "Suppressing disease spreading by using information diffusion on multiplex networks". In: *Scientific reports* 6 (2016).

[114] Amit Goyal, Wei Lu, and Laks V S Lakshmanan. "Simpath: An efficient algorithm for influence maximization under the linear threshold model". In: *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE. 2011, pp. 211–220.

[115] David Kempe, Jon Kleinberg, and Éva Tardos. "Maximizing the spread of influence through a social network". In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2003, pp. 137–146.

[116] Ming Lu et al. "TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs". In: *BMC bioinformatics* 11.1 (2010), p. 419.

[117] Mohammed Mamdani et al. "Integrating mRNA and miRNA weighted gene co-expression networks with eQTLs in the nucleus accumbens of subjects with alcohol dependence". In: *PloS one* 10.9 (2015), e0137671.

[118] Igor Ponomarev. "Epigenetic control of gene expression in the alcoholic brain". In: *Alcohol Res* 35.1 (2013), pp. 69–76.

[119] Pratheesh Sathyan, Honey B Golden, and Rajesh C Miranda. "Competing interactions between micro-RNAs determine neural progenitor survival and proliferation after ethanol exposure: evidence from an ex vivo model of the fetal cerebral cortical neuroepithelium". In: *Journal of Neuroscience* 27.32 (2007), pp. 8546–8557.

[120] Lin-Lin Wang et al. "Ethanol exposure induces differential microRNA and target gene expression and teratogenic effects which can be suppressed by folic acid supplementation". In: *Human Reproduction* 24.3 (2009), pp. 562–579.

[121] Sanjay Yadav et al. "miR-497 and miR-302b regulate ethanol-induced neuronal cell death through BCL2 protein and cyclin D2". In: *Journal of Biological Chemistry* 286.43 (2011), pp. 37347–37357.

[122] Joanne M Lewohl et al. "Up-Regulation of MicroRNAs in Brain of Human Alcoholics". In: *Alcoholism: Clinical and Experimental Research* 35.11 (2011), pp. 1928–1937.

[123] Yury O Nunez and R Dayne Mayfield. "Understanding alcoholism through microRNA signatures in brains of human alcoholics". In: *non-coding RNA and addiction* (2012), p. 17.

[124] Matthew E Ritchie et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic acids research* (2015), gkv007.

[125] Joseph J Nalluri et al. "miRsig: a consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures". In: *Scientific reports* 7 (2017).

[126] Michael Cariaso and Greg Lennon. "SNPedia: a wiki supporting personal genome annotation, interpretation and analysis". In: *Nucleic acids research* 40.D1 (2011), pp. D1308–D1312.

[127] Danielle Welter et al. "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations". In: *Nucleic acids research* 42.D1 (2013), pp. D1001–D1006.

[128] *23andMe*. URL: `https://www.23andme.com/`.

[129] *Ancestry*. URL: `https://www.ancestry.com/`.

[130] *FamilyTreeDNA*. URL: `https://www.familytreedna.com/`.

[131] Chenxing Liu et al. "MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs". In: *BMC genomics* 13.1 (2012), p. 661.

[132] Stephen T Sherry et al. "dbSNP: the NCBI database of genetic variation". In: *Nucleic acids research* 29.1 (2001), pp. 308–311.

[133] Pratip Rana et al. “Capacity estimates of additive inverse Gaussian molecular channels with relay characteristics”. In: *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies*. BICT 2015. 2015, pp. 237–240.

[134] Preetam Ghosh et al. “An analytical model to estimate the time taken for cytoplasmic reactions for stochastic simulation of complex biological systems”. In: *Proceedings of the 2006 IEEE International Conference on Granular Computing*. GrC ’06. 2006, pp. 79–84. DOI: 10.1109/GRC.2006.1635762.

[135] Albert-Laszlo Barabasi and Reka Albert. “Emergence of scaling in random networks”. In: *Science* 286 (1999), pp. 509–512.

[136] Michael Mayo et al. “Motif participation by genes in e. coli transcriptional networks”. In: *Front. Physiol.* 3 (2012), p. 357.

[137] Samik Ghosh et al. “GaMa : An Evolutionary Algorithmic Approach for the Design of Mesh-Based Radio Access Networks”. In: *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on*. IEEE. 2005, pp. 374–381.

[138] Preetam Ghosh et al. “Principles of genomic robustness inspire fault-tolerant WSN topologies: a network science based case study”. In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE. 2011, pp. 160–165.

[139] Bhanu K Kamapantula et al. “Leveraging the robustness of genetic networks: a case study on bio-inspired wireless sensor network topologies”. In: *Journal of Ambient Intelligence and Humanized Computing* 5.3 (2014), pp. 323–339.

[140] Ahmed Abdelzaher et al. “Contribution of canonical feed-forward loop motifs on the fault-tolerance and information transport efficiency of transcriptional

regulatory networks". In: *Nano Communication Networks* 6.3 (2015), pp. 133–144.

[141] Bhanu K Kamapantula et al. "Performance of wireless sensor topologies inspired by E. coli genetic networks". In: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*. IEEE. 2012, pp. 302–307.

[142] Bhanu K. Kamapantula et al. "The Structural Role of Feed-Forward Loop Motif in Transcriptional Regulatory Networks". In: *Mobile Networks and Applications* 21.1 (2016), pp. 191–205.

[143] Michael Mayo et al. "Top-level dynamics and the regulated gene response of feed-forward loop transcriptional motifs". In: *Physical Review E* 90.3 (2014), p. 032706.

[144] Michael A Rowland et al. "Crosstalk and the Dynamical Modularity of Feed-Forward Loops in Transcriptional Regulatory Networks". In: *Biophysical Journal* 112.8 (2017), pp. 1539–1550.

[145] Khajamoinuddin Syed et al. "Similar Feed-Forward Loop Crosstalk Patterns may Impact Robust Information Transport across E. coli and S. Cerevisiae Transcriptional Networks". In: *Mobile Networks and Applications* (2017).

[146] Azade Nazi et al. "Efficient Communications in Wireless Sensor Networks Based on Biological Robustness". In: *Proceedings of the 2016 International Conference on Distributed Computing in Sensor Systems*. DCOSS '16. Washington DC, 2016, pp. 161–168.

[147] Azade Nazi et al. “Exploiting gene regulatory networks for robust wireless sensor networking”. In: *Proceedings of the 2015 IEEE Global Communications Conference*. Globecom ’15. San Diego, 2015, pp. 1–7.

[148] Azade Nazi et al. “Robust deployment of wireless sensor networks using gene regulatory networks”. In: *Proceedings of the 2013 International Conference on Distributed Computing and Networking*. ICDCN ’13. San Diego, 2013, pp. 192–207. DOI: 10.1007/978-3-642-35668-1_14.

[149] Azade Nazi et al. “Deployment of robust wireless sensor networks using gene regulatory networks: An isomorphism-based approach”. In: *Pervasive and Mobile Computing* 13 (2014), pp. 246–257.

[150] Bhanu K Kamapantula et al. “Quantifying Robustness in Biological Networks Using NS-2”. In: *Modeling, Methodologies and Tools for Molecular and Nano-scale Communications* 9 (2017), pp. 273–290.