Methods 124 (2017) 78–88

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

miRTarVis+: Web-based interactive visual analytics tool for microRNA target predictions

Sehi L'Yi [a], Daekyoung Jung [a,1], Minsik Oh [a], Bohyoung Kim [b,*], Robert J. Freishtat [c,d,e], Mamta Giri [d], Eric Hoffman [f], Jinwook Seo [a,*]

[a] Department of Computer Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea
[b] Division of Biomedical Engineering, Hankuk University of Foreign Studies, 81 Oedae-ro, Mohyeon-myeon, Cheoin-gu, Yongin-si, Gyeonggi-do 449-791, Republic of Korea
[c] Division of Emergency Medicine, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010, USA
[d] Center for Genetic Medicine Research, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010, USA
[e] Department of Integrative Systems Biology, George Washington University, 111 Michigan Avenue, NW, Washington, DC 20010, USA
[f] School of Pharmacy and Pharmaceutical Sciences, Binghamton University, 4400 Vestal Parkway East, Binghamton, NY 13902, USA

ARTICLE INFO

Article history:
Received 1 April 2017
Received in revised form 2 May 2017
Accepted 2 June 2017
Available online 7 June 2017

Keywords:
microRNA
mRNA
Visualization
Expression profile
Target prediction
Integrative analysis

ABSTRACT

In this paper, we present miRTarVis+, a Web-based interactive visual analytics tool for miRNA target predictions and integrative analyses of multiple prediction results. Various microRNA (miRNA) target prediction algorithms have been developed to improve sequence-based miRNA target prediction by exploiting miRNA-mRNA expression profile data. There are also a few analytics tools to help researchers predict targets of miRNAs. However, there still is a need for improving the performance for miRNA prediction algorithms and more importantly for interactive visualization tools for an integrative analysis of multiple prediction results. miRTarVis+ has an intuitive interface to support the analysis pipeline of *load*, *filter*, *predict*, and *visualize*. It can predict targets of miRNA by adopting Bayesian inference and maximal information-based nonparametric exploration (MINE) analyses as well as conventional correlation and mutual information analyses. miRTarVis+ supports an integrative analysis of multiple prediction results by providing an overview of multiple prediction results and then allowing users to examine a selected miRNA-mRNA network in an interactive treemap and node-link diagram. To evaluate the effectiveness of miRTarVis+, we conducted two case studies using miRNA-mRNA expression profile data of asthma and breast cancer patients and demonstrated that miRTarVis+ helps users more comprehensively analyze targets of miRNA from miRNA-mRNA expression profile data. miRTarVis+ is available at http://hcil.snu.ac.kr/research/mirtarvisplus.

1. Introduction

All organisms use selective gene transcription of mRNAs to carry out biological functions. Increasingly, it is recognized that regulatory RNAs (microRNA [miRNA], long non-coding RNA [lnRNA]) play key roles in regulating the stability and translation of existing pools of mRNAs in any cell. Thus, understanding the regulation of the genome requires the integration of mRNA, as well as the regulatory RNAs targeting and regulating those mRNAs. Of the regulatory RNAs, miRNAs are the best characterized and studied. miRNAs are short, highly processed oligonucleotides (approx. 22 nt) that carry out post-transcriptional regulation of target mRNAs through either degradation of the target mRNA or inhibition of protein translation [1].

The specific mRNA targets for any specific miRNA can be derived bioinformatically through miRNA-mRNA sequence alignment and evolutionary conservation of the target mRNA sequence. For example, miRNA target prediction algorithms such as TargetScan [2] or miRanda [3] predict targets of miRNAs. The potential interactions between any miRNA-mRNA pair require experimental validation, typically through reporter constructs.

Recent prediction algorithms used miRNA-mRNA expression profile data. Microarray method has been prevalent before deep sequencing method becomes popular recently for miRNA-mRNA expression profiling. As deep sequencing methods become widespread, whole genome miRNA-mRNA expression profile data

Abbreviations: miRNA, microRNA; TCGA, The Cancer Genome Atlas.

* Corresponding authors.
E-mail addresses: bkim@hufs.ac.kr (B. Kim), jseo@snu.ac.kr (J. Seo).

[1] Present address: Samsung Electronics, South Korea.

http://dx.doi.org/10.1016/j.ymeth.2017.06.004
1046-2023/

become widely available. Accordingly, some algorithms exploited miRNA-mRNA expression profile data to search for targets of miRNAs. Bioinformaticians also introduced Web-based tools [4–7] to integrate miRNA-mRNA expression profile data with sequence-based miRNA target prediction algorithms.

However, those tools are limited in supporting rich exploratory analysis of miRNA-mRNA expression profile data. For example, enabling dynamic queries and providing relevant biological information on demand in the visualizations are among the much-expected features. We believe that more work is required in designing visualizations and interactions of visual analysis tools for miRNA-mRNA expression profile data. Given that both miRNA and mRNA expression datasets are multidimensional, searching for mRNA targets requires integrative analysis of the two heterogeneous multidimensional datasets. To obtain more improved accuracy of such integrative multidimensional data analysis, interactive visual analysis tools for miRNA-mRNA expression profile data should help researchers

- predict miRNA-target interactions by integrating miRNA-mRNA expression profile datasets and
- understand the structure of miRNA-mRNA interaction network.

This paper is an extended version of our paper for the BioVis Conference, Dublin, Ireland, 2015 [8]. In this paper, we present miRTarVis+, a Web-based interactive visual analytics tool that predicts and visualizes miRNA-target interaction network using miRNA-mRNA expression profile data. miRTarVis+ is an enhanced version of miRTarVis [8], which is a desktop Java application. We redesigned miRTarVis+ as a Web-based visual analytics system to make it run on commodity Web-browsers. Based on observations and interviews, we first defined a common analysis pipeline for miRNA-mRNA expression profile data and then designed the interface of miRTarVis+ based on the analysis pipeline. miRTarVis+ provides prediction algorithms that are based on both sequence and expression profile data. miRTarVis+ is the first visual analytics tool that applies GenMiR++ [9], a Bayesian inference model, and maximal information-based nonparametric exploration (MINE) analysis [10], a new technique that finds highly associated pairs from multidimensional data, to predict targets of miRNAs from miRNA-mRNA expression profile data.

Compared to miRTarVis [8], miRTarVis+ enables users to perform integrative analysis in the prediction results level as well. It provides a *prediction overview* where users can compare and combine multiple target prediction results and select a subset of miRNA-mRNA pairs of their interest with multiple prediction results integrated. When users select miRNA-mRNA pairs, miRTarVis+ visualizes the resulting bipartite miRNA-target regulatory network in interactive treemap and node-link diagram. The treemap is a unique feature of miRTarVis+, and it is expected to outperform a node-link diagram visualization when miRNA-target interactions are overcrowded. We report results of two case studies (including a new case study in addition to the previously introduced one [8]) to prove the efficacy of miRTarVis+ by applying it to human miRNA-mRNA expression profile data.

2. Methods

2.1. Design goals and rationales

At the first stage of our iterative design process, we tried to define a common analysis pipeline for miRNA-mRNA expression profile data. We observed researchers who analyze miRNA data and conducted informal interviews with them. We identified four major analysis steps in their analysis process for miRNA-mRNA expression data, which constitute the analysis pipeline as follows:

1. *load*: load miRNA-mRNA expression profile data
2. *filter:* filter miRNA-mRNA expression profile data to leave only significant miRNAs and mRNAs for further analysis
3. *predict:* predict miRNA-target interactions by sequence-based prediction algorithms and search for highly associated miRNA-mRNA pairs in expression profile data by data mining or machine learning techniques
4. *visualize:* visualize the resulting miRNA-target network to help researchers understand the network structure and biological implication of the network

After deriving this analysis pipeline, our long-term design collaborative design process with biomedical researchers led us to the following design goals of our visual analysis. It should help users

1. analyze miRNA-mRNA expression profile data of various types based on the analysis pipeline,
2. improve miRNA target prediction accuracy by integrating multiple target prediction algorithms, and
3. comprehend the resulting miRNA-mRNA interaction network through interactive visualizations.

To achieve these design goals, we designed and implemented our visualization tool based on the following design rationales:

1. provide a user interface based on the analysis pipeline
2. support various types of miRNA-mRNA expression profile data
3. provide interactive filtering of less significant or erroneous miRNAs and mRNAs for better prediction accuracy
4. integrate diverse prediction algorithms, including novel prediction algorithms, for more accurate prediction results
5. present analysis results in intuitive visualizations
6. support dynamic queries through intuitive user interactions to help users search biological findings

2.2. Unique features of miRTarVis+

The user interfaces of miRTarVis+ were designed based on the analysis pipeline for miRNA-mRNA expression profile data. In accordance with the analysis pipeline (i.e., *load*, *filter*, *predict*, and *visualize*), we organized the four steps of the pipeline using a step-by-step navigation interface (Fig. 1).

miRTarVis+ gives users more flexibility in preparing input miRNA-mRNA expression profile data. It can accept both two-sample and multisample miRNA-mRNA expression profile data. It also accepts data that only consists of fold change and p-value without underlying expression data. Moreover, miRTarVis+ directly accepts TCGA (The Cancer Genome Atlas) miRNA-mRNA expression profile data.

miRTarVis+ supports filtering functions most appropriate for individual input data types. For two-sample expression profile data, users can filter data by p-value and fold change of each mRNA and miRNA, which are calculated automatically on data loading. For multisample expression profile data, users can filter out poorly expressed (e.g., most of expression levels being zero) miRNAs and mRNAs.

miRTarVis+ is the first tool that applies the MINE analysis [10] to the search for targets of miRNAs from miRNA-mRNA expression profile data. The MINE analysis is adopted to support finding more general relationships because it can search for not only conventional linear relationships but also nonlinear and nonfunctional relationships. miRTarVis+ also supports finding causal relationships

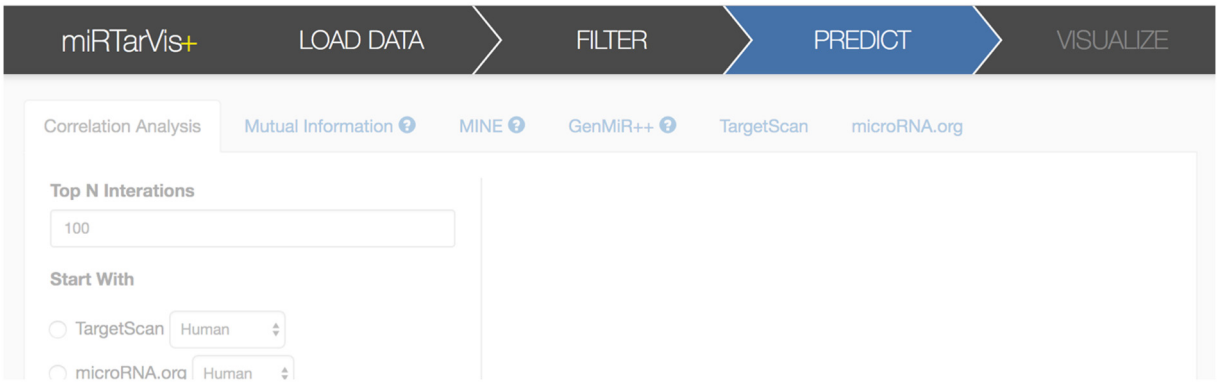

Fig. 1. A step-by-step navigation interface in miRTarVis+. We organized the four steps of the analysis pipeline (i.e., *load*, *filter*, *predict*, and *visualize*) using the navigation interface to provide good affordance that enables users to naturally follow the analysis pipeline, step by step.

between miRNAs and their targets from miRNA-mRNA expression profile data by adopting a Bayesian inference modeling analysis (GenMiR++ [9]). miRTarVis+ supports conventional correlation analysis and mutual information analysis as well.

miRTarVis+ provides more flexibility in analyzing using multiple prediction algorithms by supplying a prediction overview, where users can select a subset of their interest among all miRNA-mRNA pairs as well as see overall information of multiple target prediction results.

miRTarVis+ is the first tool that adopts a treemap to show a resulting miRNA-target regulatory network in such a way that a miRNA node encloses its target mRNAs in a treemap layout. In this way, the treemap is more space-efficient than the traditional node-link diagram when the network is overcrowded by a large number of miRNA-target interactions. miRTarVis+ also visualizes a miRNA-target network in an improved node-link diagram where multiple miRNA-target nodes are connected to a miRNA node. Users can navigate the diagram interactively for closer inspection of an interesting miRNA or mRNA. Users can also move miRNA or mRNA nodes to a better position. This enhanced interactivity can help users understand the structure of a miRNA-target network and can make a node-link diagram more suitable for publication. miRTarVis+ also helps researchers access relevant detailed information for miRNAs and mRNAs by linking a corresponding website that contains the information upon a mouse click. We will discuss this in detail later when we explain the visualizations of miRNA-target network.

2.3. UI and functions of miRTarVis+

We designed miRTarVis+ based on the *load-filter-predict-visualize* pipeline (Fig. 1). We adopted a simple step-by-step navigation interface, where only one selected menu item is opened while all others are hidden. This is more intuitive to users because the order of the menu items matches the order of the analysis procedure. Through this step-by-step interface, users can focus on the current step of the procedure while using the screen space more effectively even in a small-size screen (down to a screen resolution of 960×720px).

There are four main menus: *load data*, *filter*, *predict*, and *visualize*. In the *load data* menu (Fig. 5), miRTarVis+ can load one of five types (paired two-sample, unpaired two-sample, *p*-value and fold change, multisample, and TCGA) of miRNA-mRNA expression profile data. In the *filter* menu, miRTarVis+ can select only significant miRNAs and mRNAs that will serve as a search space in the next predict step. miRTarVis+ provides different filtering options for two-sample and multisample type data: filtering by *p*-value or fold change for the former and filtering by average expression level for the latter. miRTarVis+ shows histograms of *p*-value and fold change for two-sample type data and a histogram of average expression level for multisample type data.

In the *predict* menu (Fig. 6), miRTarVis+ searches for targets of miRNAs by multiple algorithms. The predict menu uses a tabbed interface where each tab is dedicated to a prediction algorithm. When a user selects a certain prediction algorithm in a tab, the interface for choosing parameters for the selected algorithm appears on the left. miRTarVis+ searches for miRNA-target pairs among miRNAs and mRNAs that survived from the previous *filter* step. After finishing the prediction process, miRTarVis+ shows the predicted miRNAs and their targets in a table on the right.

In the *visualize* menu, miRTarVis+ provides three interactive visualizations: *prediction overview*, *node-link diagram*, and *treemap*. In the prediction overview, users can aggregate and select a subset of the prediction results of their interest derived from different prediction algorithms (e.g., select commonly predicted miRNA-mRNA interactions in all results). The selected miRNA-target interactions are then visualized with the node-link diagram and treemap (Figs. 3 and 4). In both visualizations, miRTarVis+ provides external links to miRBase [11] and miR2Disease [12] for miRNAs and to NCBI and GeneCards [13] for mRNAs through a context menu. miRTarVis+ also provides gene enrichment analysis (provided by the Gene Ontology Consortium web service) of target genes of a miRNA. With accurate prediction, the biological functions of predicted targets of a miRNA tend to be similar. Through enrichment analysis, miRTarVis+ provides a convenient way for confirming the function of a miRNA and validating the target prediction result.

When users cannot find satisfactory prediction results in the *visualize* menu, they can go back to the *filter* or *predict* steps to change the parameters of filtering or prediction for a better result. Through this iterative procedure, miRTarVis+ can help users narrow down to more important and interesting miRNA-target interactions.

2.4. Prediction of miRNA targets

In the *predict* step, miRTarVis+ searches for miRNA-mRNA target interactions among all remaining miRNA-mRNA pairs from the previous step (i.e., *filter* step). miRTarVis+ uses prediction algorithms that are based on both sequence and expression profile data. miRTarVis+ supports two sequence-based prediction algorithms, TargetScan and microRNA.org, which are two of the most cited miRNA target prediction algorithms. miRTarVis+ supports four techniques for prediction algorithms based on expression pro-

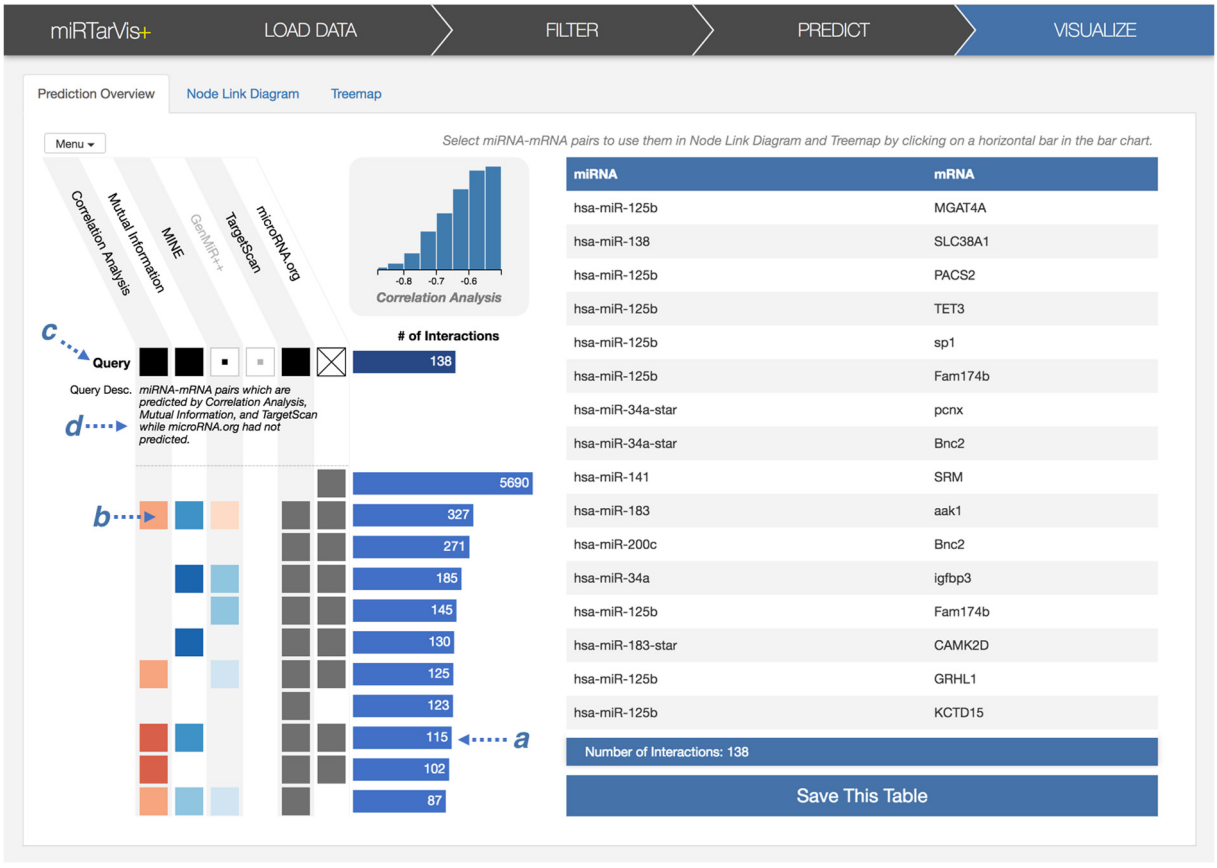

Fig. 2. Overview of multiple prediction results. miRTarVis+ provides overall information about prediction results of different target prediction algorithms.

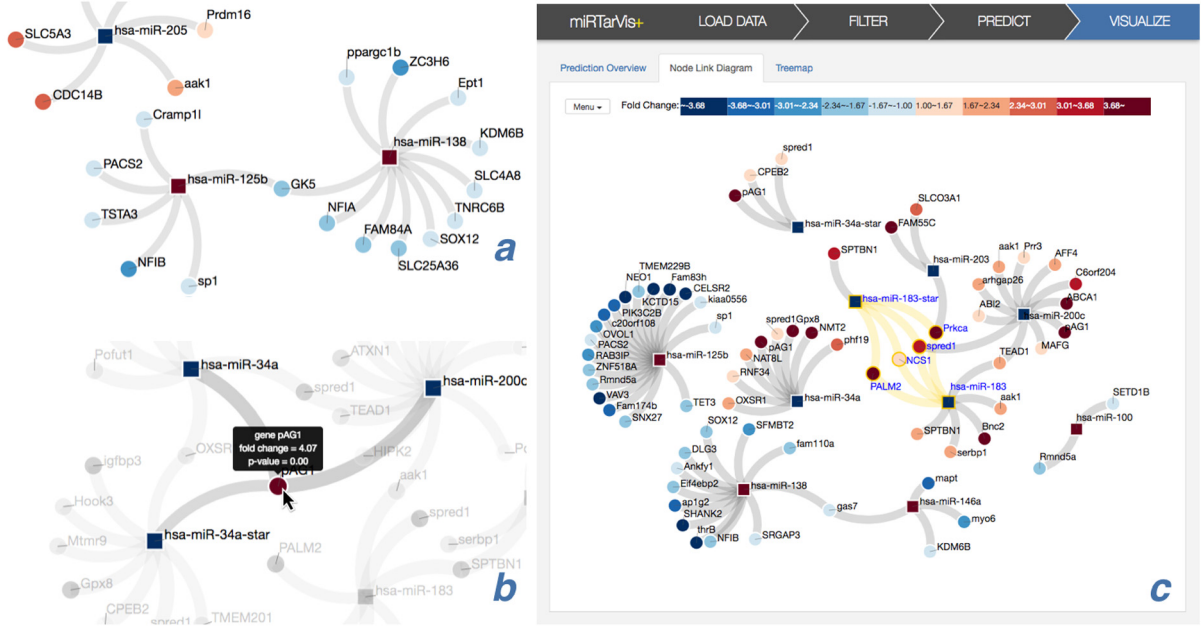

Fig. 3. A node-link diagram visualization in miRTarVis+. It visualizes a miRNA-mRNA regulatory network. Overall topology of the network is more clearly shown in the node-link diagram.

file data, and these techniques are correlation analysis (a user can choose to use Pearson or Spearman coefficient), mutual information analysis, Bayesian inference model analysis (GenMiR++), and MINE analysis. Given that prediction algorithms based on expression profile data require expression level at each sample, miRTarVis+ does not support this type of prediction for the third data type of *p*-value and fold change.

Prediction algorithms based on expression profile data calculate scores that represent the intensity of association for each miRNA-target pair. miRTarVis+ allows users to set a threshold value to fil-

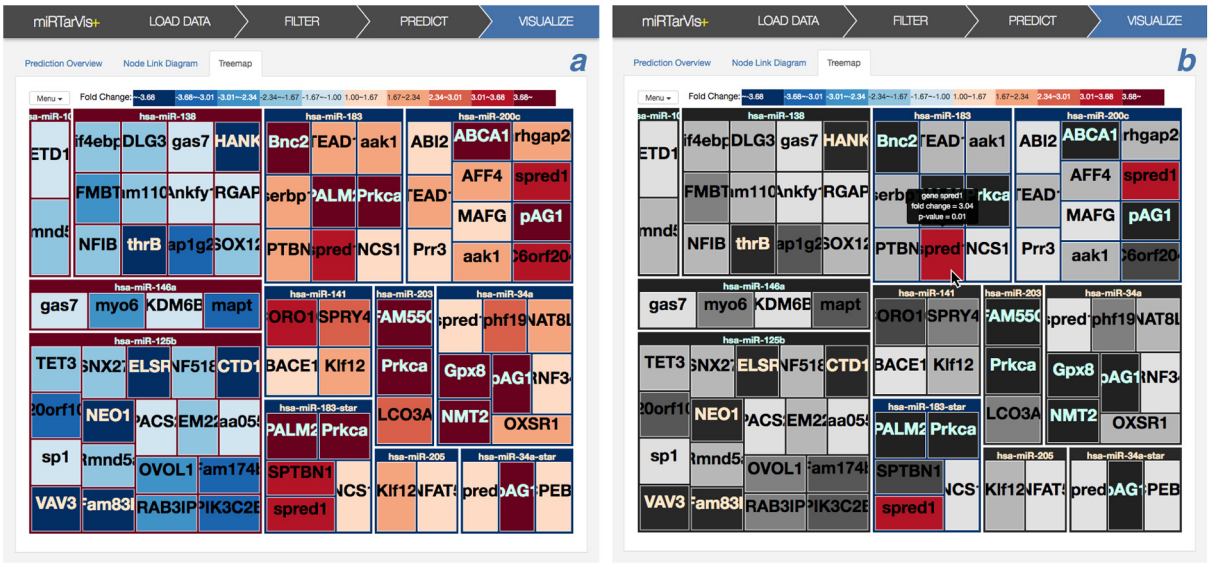

Fig. 4. A treemap visualization in miRTarVis+. It visualizes a miRNA-mRNA regulatory network. Red and blue represent up- and down-regulated fold changes, respectively. Color saturation represents the intensity of a fold change.

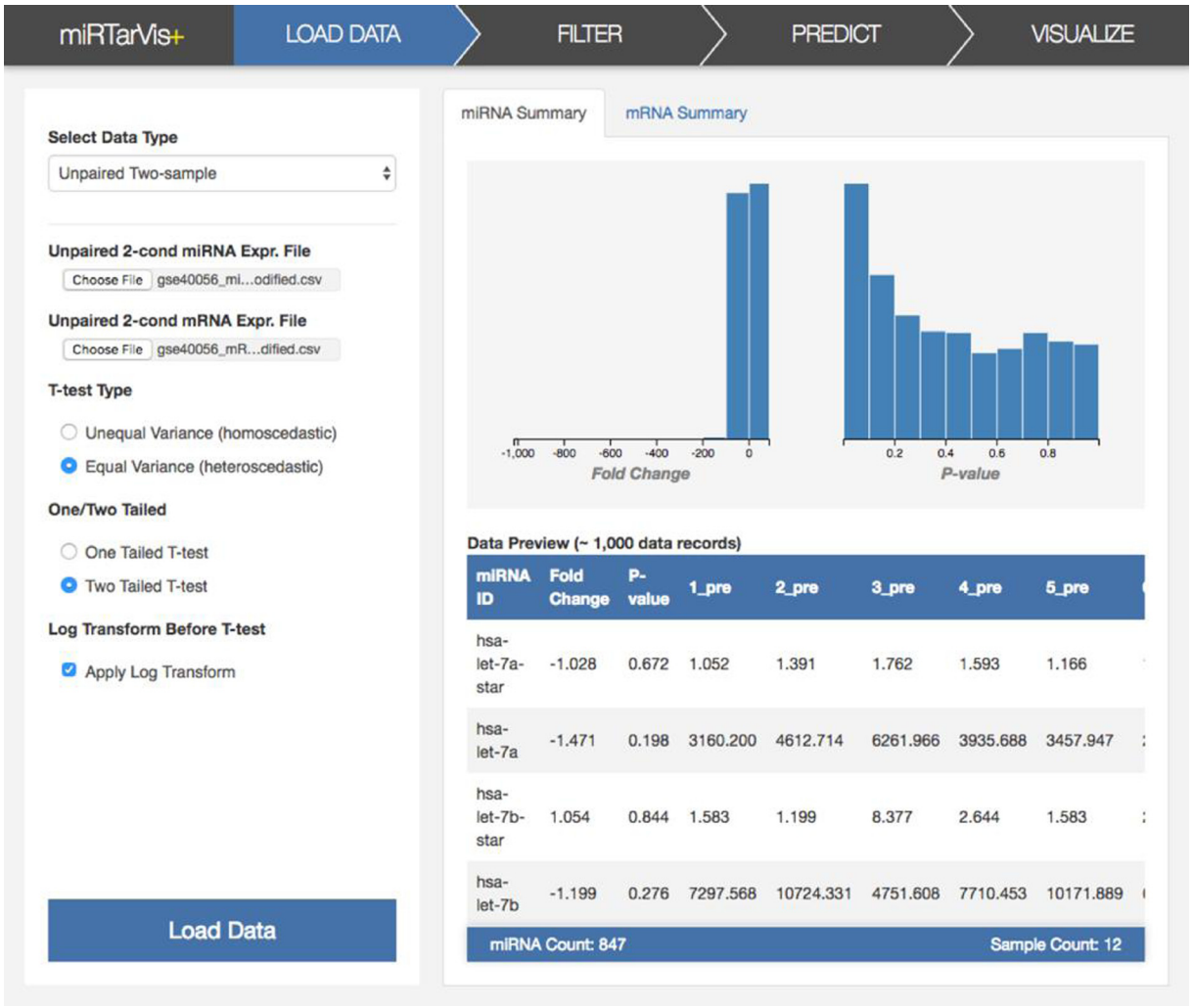

Fig. 5. Data load menu in miRTarVis+. miRTarVis+ can load one of five types of miRNA-mRNA expression profile data, including paired two-sample, unpaired two-sample, p-value and fold change, multisample, and TCGA.

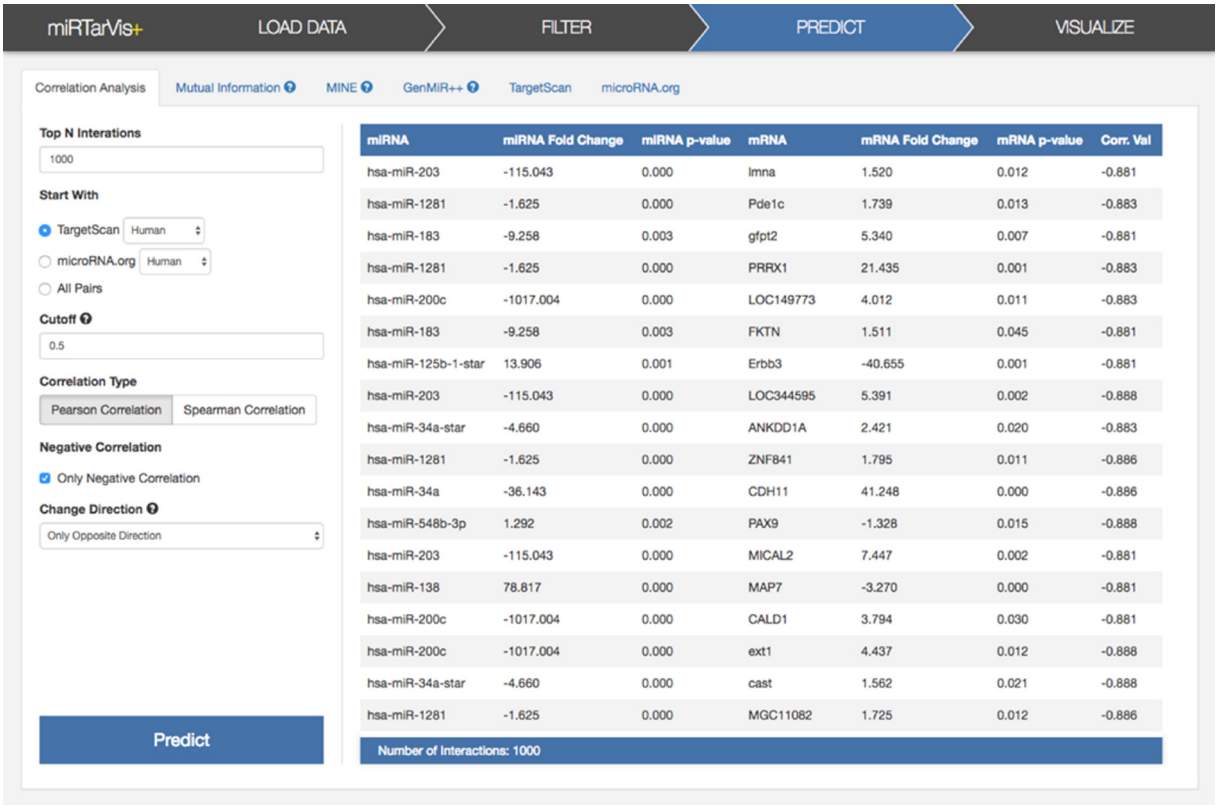

Fig. 6. Prediction menu in miRTarVis+. miRTarVis+ predicts targets of miRNA by adopting Bayesian inference and MINE analyses, as well as conventional correlation and mutual information analyses.

ter miRNA-mRNA interactions by their score when conducting the predictions. miRTarVis+ can also set the number of resulting miRNA-target interactions for prediction. In this case, miRTarVis+ searches for the top high-scored miRNA-target interactions.

miRTarVis+ can filter miRNA-mRNA interactions by their fold change direction as well. For example, predicted miRNA-mRNA interactions where miRNA is up-regulated and mRNA is down-regulated are biologically more significant because miRNAs down-regulate their target mRNAs. To enable diverse filtering by fold change direction, miRTarVis+ provides four search options for fold change direction of miRNA-mRNA interactions in prediction: up-regulated miRNA and down-regulated mRNA, down-regulated miRNA and up-regulated mRNA, oppositely regulated (union of the first and the second options), and all pairs. If prediction is confined to miRNA-mRNA pairs that consist of up-regulated miRNA and down-regulated mRNAs or pairs that consist of down-regulated miRNA and up-regulated mRNAs, the accuracy of prediction could be improved. Therefore, miRTarVis+ enables users to use these options to improve the accuracy of their miRNA target prediction.

2.5. Visualization: Overview of multiple target prediction results

In the *visualize* step, miRTarVis+ generates a regulatory network for miRNA-mRNA pairs. To support an integrative analysis of multiple prediction results, miRTarVis+ first provides the overall information of the multiple prediction results (e.g., common miRNA-mRNA interactions between results) in the *prediction overview* (Fig. 2). Users can select a subset of miRNA-mRNA interactions from different prediction algorithms in the prediction overview, and then a regulatory network of the selected subsets is visualized (explained in detail in the next section).

The prediction overview is newly introduced feature in miRTarVis+. The prediction overview adopts UpSet [14], which is a matrix-based visualization technique for quantitative analysis of sets. In miRTarVis+, each column in the overview represents a prediction algorithm while each row represents an exclusive subset of miRNA-mRNA pairs among all results. In each cell, a rectangle is placed only when the corresponding subset of miRNA-mRNA pairs (row) is predicted by the corresponding prediction algorithm (column). Unlike UpSet which uses gray-scaled circles for filling cells, we color-encode the rectangles (Fig. 2b) to show the prediction score (e.g., correlation coefficient for the correlation analysis) of the corresponding subset when the corresponding prediction algorithm gives any score. The same ten-level discrete color scheme as in representing the fold changes (explained later) shown in Fig. 3c is used. When a subset of miRNA-mRNA pairs is predicted by a prediction algorithm but the algorithm does not give score, the corresponding cell is filled with a dark-gray rectangle. Using the rectangular shape instead of the conventional circular one is to more clearly show the color of each cell by rendering color with larger area, which is also a familiar way of visualizing subsets as in ConSet [15].

At the end of each row, a bar is placed to show the number of miRNA-mRNA pairs in the corresponding subset. For example, in Fig. 2a, the row indicated by a blue arrow represents that there are 115 miRNA-mRNA pairs that are commonly predicted by correlation, mutation information, TargetScan, and microRNA.org analyses. Since bars could be too small if they contain relatively small number of miRNA-mRNA pairs, users can apply log transformation to the length of all bars from the popup menu on the top-left corner of the prediction overview.

miRTarVis+ enables users to query for miRNA-mRNA pairs of their interest. The rectangles above the first row of the subsets of miRNA-mRNA pairs are for composing a query. Three kinds of a

rectangular symbol, ■, ☒, and ■ are for inclusion, exclusion, and no consideration of a given subset, respectively (Fig. 2c). A black-filled rectangle represents that the miRNA-mRNA pairs predicted by the corresponding prediction algorithm are included in the query results; a rectangle surrounding an X mark represents that the miRNA-mRNA pairs predicted by the corresponding algorithms are excluded from the query results; and a small rectangle represents that the corresponding algorithm is not used as an inclusion or exclusion criterion, thus the corresponding miRNA-mRNA pairs may or may not be included in the results. For example, the query in Fig. 2c is for finding miRNA-mRNA pairs which are commonly predicted by correlation, mutation information, and TargetScan analyses but which are not predicted by the microRNA.org analysis. To help users better understand the meaning of a current query, we additionally placed a simple natural language description below the query (Fig. 2d).

2.6. Visualization: Visualizations of miRNA-target network and interactions

After users select miRNA-mRNA pairs from the *prediction overview*, miRTarVis+ generates a regulatory network using the selected miRNA-mRNA pairs and visualizes a resulting bipartite graph both in a node-link diagram and a treemap.

To achieve the aforementioned design goals of miRTarVis+, visualizations and interactions in miRTarVis+ should be designed to help users effectively understand the predicted miRNA-target network. More specifically, miRTarVis+ should provide the following functionalities:

- reveal topological structures in the network
- highlight mRNAs that are regulated commonly by multiple miRNAs in the network
- provide information of particular miRNAs and mRNAs on demand
- enable users to go back to the previous step and adjust parameters to update the miRNA-target interaction network

We adopt a node-link diagram and a treemap visualization to achieve the design goals. The node-link diagram is a more familiar visualization in bioinformatics research fields [5,6,16] which enabled miRTarVis+ to help reveal the overall topological structures of a miRNA-target interaction network, while the treemap visualization can better reveal mRNA targets for each miRNA in a more space-efficient manner without occlusion.

In the node-link diagram (Fig. 3), each node represents a miRNA or mRNA and each link represents the interaction between the miRNA and mRNA. miRNAs and mRNAs are visually distinguished using different shapes (i.e., rectangular nodes for miRNAs and circular ones for mRNAs). The width of each link represents the number of prediction algorithms that predicted the corresponding miRNA-mRNA interaction. We used a force-directed layout to delineate the shape of the miRNA-target interaction network by using one of the most popular visualization library, D3.js (http://d3js.org). Repulsive and attractive force strengths, which are two important hyperparameters of the force-directed layout that affect the position of nodes, were empirically determined to more clearly show the overall topology of the network with least visual occlusion between nodes, labels, and links. In addition, miRTarVis+ enables users to place node labels (i.e., names of miRNAs and mRNAs) without overlaying other labels or nodes by using a simulated annealing based plug-in [17] (Fig. 3a). Users can use this feature from a popup menu on the top-left corner of the diagram.

However, the node-link diagram still has an innate limitation of visual occlusion when data having excessive nodes and links. To mitigate this limitation, we additionally provided various user interactions. Users can relocate a node manually to avoid the occlusion. When users move a mouse cursor on a node, miRTarVis+ highlights the nodes and links which are directly connected to the node while dimming other nodes and links (Fig. 3b). When users select a node by left click on the node, miRTarVis+ highlights the connected links and nodes in yellow. If users select multiple nodes, miRTarVis+ highlights the commonly linked nodes to those selected nodes in yellow. For example, in Fig. 3c, a user clicked on two miRNAs (i.e., has-miR-183-star and has-miR-183), and then all their common target mRNAs (i.e., PALM2, NCS1, spred1, and Prkca) are highlighted in yellow. This feature is valuable because commonly targeted mRNAs could play an important role in a miRNA-target network. Users can clear the node selection by clicking on a background of the node-link diagram. In addition, users can navigate (i.e., zoom and pan) the miRNA-target network using simple mouse wheel and drag interactions.

miRTarVis+ can also visualize a miRNA-target regulatory network using a treemap (Fig. 4). One characteristic of miRNA-target interaction networks is that one miRNA node is connected to many target mRNA nodes. We could convert the network into a two-level hierarchy where a miRNA is a parent of target mRNAs by exploiting this characteristic, and a treemap visualization can be used to represent the network. In the tree, leaf nodes represent mRNA nodes, and their parent nodes represent miRNA nodes that have link with the mRNAs. Therefore, if one mRNA node has links with multiple miRNA nodes, the mRNA node occurs multiple times in the treemap visualization. In the original treemap, the area encodes an attribute of the data. However, in our treemap visualization, all mRNA nodes have the same area. Therefore, the size of a miRNA node reflects the number of its target mRNAs in the network. In addition, font size of each mRNA is varied depending on the number of prediction algorithms which predicted the corresponding miRNA-target pair.

Compared with a node-link diagram, a treemap visualization can represent a complex regulatory network without occlusion among links and nodes, especially when there are too many links crossing each other in a node-link diagram. Given that a treemap visualization is much more space-efficient than a node-link diagram, more screen space can be devoted to showing gene symbol names without occlusion; these names serve as important information for biologists to understand the biological functions of mRNAs in a miRNA-target interaction network. For example, in Figs. 3c and 4a, a miRNA-target interaction network is visualized in both treemap and node-link diagram. Identifying targets of a miRNA of the node-link diagram is more difficult than in treemap because there is not much room for showing gene symbols for target mRNAs without occlusion.

However, the treemap visualization has a disadvantage of not showing the overall structure of the network. miRTarVis+ represents all miRNAs as top-level nodes and all their targets as their child nodes. As a result, if multiple miRNAs have a common target mRNA, the mRNA node appears multiple times in the treemap, and there is no affordance to imply whether an mRNA is connected to multiple miRNAs. We tried to resolve this problem by interactively highlighting all nodes representing a selected mRNA node in our treemap visualization (Fig. 4b).

miRTarVis+ uses color as a unique visual cue to help users easily recognize whether a predicted miRNA-mRNA interaction is supported by input experimental miRNA-mRNA expression profile data. For two-sample data, the fold change value is color coded: down-regulated miRNAs and mRNAs in blue and up-regulated miRNAs and mRNAs in red (Fig. 3 and 4). The intensity of fold change is represented by color saturation, and darker color means higher degree of fold change. Therefore, users can easily grasp whether a given prediction is supported by the input data or not, by comparing the colors of the ends of a link. For color-blind users,

a different color scheme from ColorBrewer (http://colorbrewer2.org) can be selected from a popup menu on the top-left corner of each visualization. For multi-sample miRNA-mRNA expression profile data, miRTarVis+ represents miRNAs in orange and mRNAs in dark blue.

2.7. Implementation

miRTarVis+ was mainly implemented using HTML, CSS, and JavaScript, which enables our system to run on any platforms with regular web browsers. To implement visualizations, we used D3.js (ver. 4.7.4, http://d3js.org) which is one of the most popular visualization library for JavaScript. To implement prediction algorithms such as mutual information estimation [18], GenMiR++ [9], and MINE [10] analyses, we used pure JavaScript and math.js (ver. 3.10.3, http://mathjs.org). We embedded a TargetScan [2] database into miRTarVis+. We downloaded a miRNA family table and a predicted conserved target information table, and we joined the two tables into one. The resulting table contains three attributes: miRBase ID, gene symbol, and species. miRTarVis+ has the resulting table that contains a set of conserved targets of all miRNAs for nine species (human, mouse, rat, rhesus, frog, dog, cow, chimpanzee, and chicken). miRTarVis+ also embedded databases from microRNA.org [19] (http://www.microrna.org). We downloaded target predictions with "Good mirSVR Scores" and "Conserved miRNA" from the August 2010 release from the website and embedded it into miRTarVis+.

3. Results and discussion

To evaluate the efficacy of miRTarVis+, we performed two case studies using miRNA-mRNA expression profile data of asthma and breast cancer patients.

3.1. Case study 1: Applying to miRNA-mRNA expression profile data of asthma patients

We conducted the first case study with a group of researchers (led by the fifth author) studying childhood asthma in Washington DC area. As the rate of comorbid asthma and obesity increases, identifying mechanisms by which obesity affects asthma is critical. The group reported that obese visceral adipocytes shed exosomes containing miRNAs that can up-regulate the expression of profibrotic signaling genes in the lung [20]. An important next step in their analyses was to define the set of lung mRNA responses to these adipocyte-derived exosomes. Prior standard approaches would have included generating a list of potential target mRNAs and prioritizing them for validation. miRTarVis+ presented the researchers with a new opportunity to objectively define their target validation set of mRNAs using multiple in silico analyses.

The participating researchers used miRTarVis+ for a couple of hours after a short tutorial session using remote video conferencing. Using airway fibroblasts (i.e., cells important in the development of lung fibrosis), they demonstrated the use of miRTarVis+ to define potential target mRNAs through which obesity can induce lung fibrosis in asthma. They used obese visceral adipocyte-derived exosomes (n = 4) that were previously tested for miRNA expression (Affymetrix microRNA 3.0 array). They coincubated these exosomes with human airway fibroblasts (from endobronchial biopsy tissue) from nonasthmatic and asthmatic donors (n = 1 each) for 24 h. Fibroblasts were profiled for global mRNA expression.

One of the major advantages of miRTarVis+ is that it enables instantaneous application of multiple analytical algorithms to the data. Normalized, background-subtracted miRNA-mRNA expression profile data were imported into miRTarVis+ and filtered for a paired two-tailed t-test with $p \leq 0.05$. Multiple prediction algorithms (i.e., Pearson correlation, MINE, GenMiR++, and TargetScan) were applied, and the top 1000 negative correlations and top 100 opposite change directions were selected in each algorithm. The intersection among them could be identified through the prediction overview in miRTarVis+.

They identified 45 miRNA-mRNA pairs (15 miRNAs/33 mRNAs) for obese visceral exosomes and nonasthmatic fibroblasts and 61 miRNA-mRNA pairs (33 miRNAs/27 mRNAs) for asthmatic fibroblasts. Their focus turned to ACVR2B (activin receptor, type IIB; myostatin and TGFβ receptor) as the only gene present in both datasets, that is, down-regulated in nonasthmatic fibroblasts (fold change [FC] = −1.18, $p < 0.01$) and up-regulated in asthmatic fibroblasts (FC = 1.31, $p = 0.02$). Fig. 7 shows that obese visceral exosomal miRNAs targeting ACVR2B were up-regulated in nonasthmatic fibroblasts (i.e., hsa-let-7b-star_st [FC = 2.31, $p = 0.027$] and hp_hsa-mir-3118-5_x_st [FC = 2.16, $p = 0.025$]) and down-regulated in asthmatic fibroblasts (hp_hsa-mir-103a-1_st [FC = −2.47, $p < 0.001$], hp_hsa-mir-103a-1_x_st [FC = -2.18, $p = 0.003$], hp_hsa-mir-23a_x_st [FC = −1.08, $p = 0.035$], hp_hsa-mir-3118-1_x_st [FC = −1.53, $p = 0.029$], hp_hsa-mir-3118-6_x_st [FC = −1.56, $p = 0.008$], hp_hsa-mir-320b-1_st [FC = −2.41, $p = 0.002$], hp_hsa-mir-320b-2_st [FC = −1.51, $p = 0.019$], and hp_hsa-mir-320c-1_x_st [FC = -2.37, $p = 0.004$]). qRT-PCR confirmed ACVR2B down-regulation in nonasthmatic fibroblasts (FC = 0.26, 95% confidence interval = [0.26, 0.78]) and up-regulation in asthmatic fibroblasts (FC = 3.21, [3.21, 6.72]).

In summary, the participating researchers said that they were able to quickly and inexpensively identify a biologically relevant mRNA target for adipocyte-derived exosomal miRNAs. This target, ACVR2B, is down-regulated in nonasthmatic fibroblasts and up-regulated in asthmatic fibroblasts, suggesting that obese visceral adipocyte-derived exosomes regulate airway fibroblast gene expression and that these cells respond differently to the exosomes depending on disease state. miRTarVis+ enabled this novel mechanistic discovery by which adiposity may increase lung fibrosis in asthma.

3.2. Case study 2: Finding miRNA-mRNA pairs that play important role in breast cancer

We conducted a second case study with a bioinformatician (third author of this paper) who has two years of experience in miRNA target prediction analysis. Previously, he usually used magia2 [6], cosmic [21], and MMIA [4] for the analysis, and it was his first time using miRTarVis+ for his research. Pre- and post-study interviews were conducted each for ten minutes. The participant used miRTarVis+ for 90 min after a short tutorial session (15 min long). He used a dataset containing microarray expression profiles of 847 miRNAs and 20,076 mRNAs of four invasive breast cancer cell lines (MDA-MB-231, HS578T, BT549, SUM159), six less invasive cell lines (BT474, MDA-MB-468, T47D, ZR-75-1, MCF7, SK-BR3), and two non-tumorigenic breast epithelial cell lines (MCF10A and MCF12A) [22]. The dataset had been prepared before the study so the participant was able to focus only on the analysis during the study.

The participant's task was to find miRNAs and their target mRNAs that play an important role in breast cancer cell invasion. For the analysis, he first loaded the dataset to miRTarVis+ and performed unpaired t-test between two groups (i.e., eight less invasive or non-tumorigenic cell lines and four invasive cell lines). Then, he filtered out miRNAs and mRNAs whose p-value and |fold change| were less than 0.05 and 1.2, respectively. He then performed a single prediction analysis (i.e., Pearson correlation) to first see the overall interaction trend of the dataset. Top 300 negatively correlated miRNA-mRNA pairs were visualized in node-link diagram

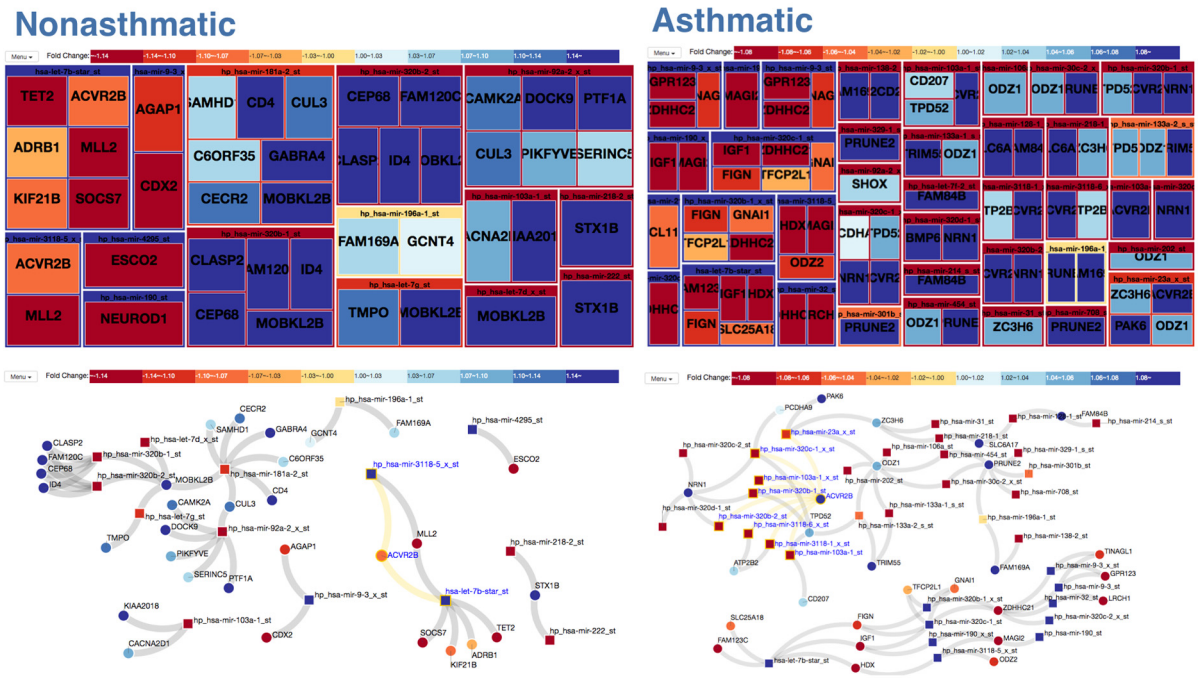

Fig. 7. Results of the first case study. Interrelations among miRNAs and mRNAs for nonasthmatic or asthmatic fibroblasts exposed to obese visceral exosomes.

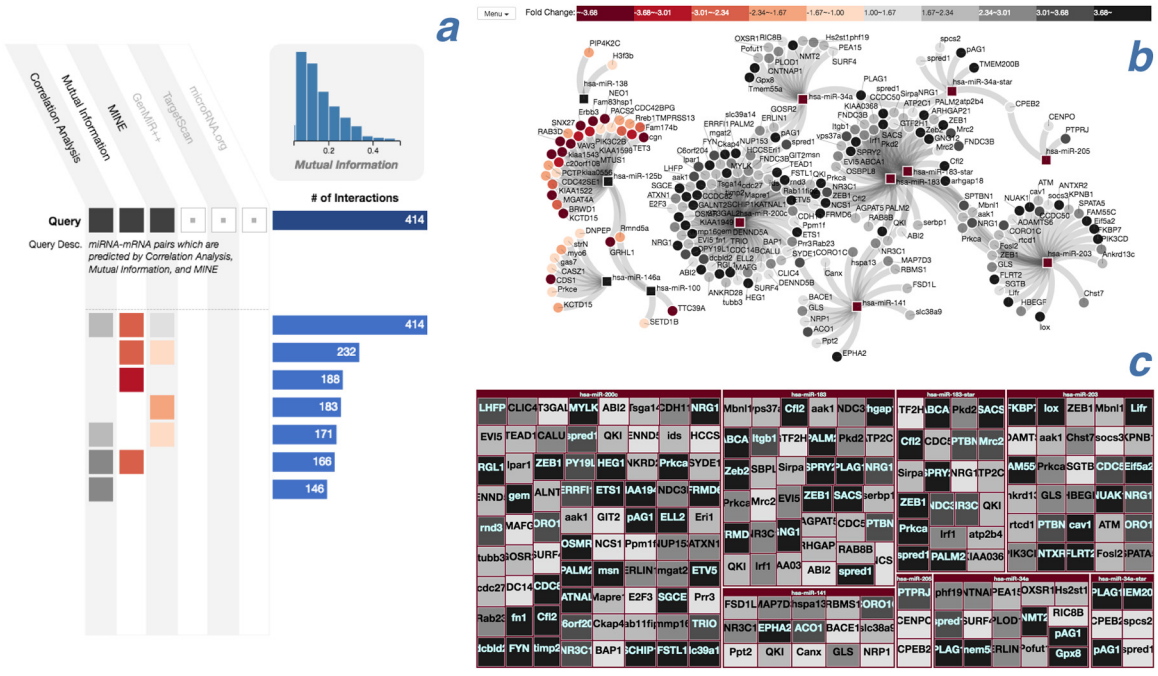

Fig. 8. Results of the second case study. Many mRNAs which are overexpressed in invasive breast cancer cell lines are targeted by the miR-200c which was significantly under-regulated in invasive breast cancer cell lines.

and treemap visualizations. After examining the visualizations, he concluded that the prediction result contained too many miRNAs, which makes the prediction algorithms produce too many miRNA-mRNA pairs. So, he went back to *filter* step and reduced the number of miRNAs by lowering the *p*-value threshold to 0.005. After the filtering, only 20 miRNAs and 6028 mRNAs remained.

Then, he performed multiple prediction analyses (i.e., Pearson correlation, mutual information, and MINE), and selected top 300 miRNA-mRNA interactions of opposite change directions for each algorithm. The prediction overview showed interaction subsets among three prediction results. In the visualization, he could easily notice that only a few interactions were commonly predicted between pairs of three algorithms (i.e., 32 interactions between

correlation analysis and mutual information, five interactions between mutual information and MINE, and other five interactions between correlation analysis and MINE), and no interactions were commonly predicted in all three algorithms. He was not satisfied with the result since he wanted to find miRNA-mRNA pairs that were stably predicted by three algorithms (i.e., Pearson correlation, mutual information, and MINE) with which he was familiar. So, he selected 700 more interactions for each analysis (i.e., top 1000 interactions) when performing the integrative prediction analysis again. As a result, the prediction overview showed that 414 interactions were commonly predicted by three algorithms (Fig. 8a).

He selected these common interactions to visualize them in node-link diagram and treemap visualizations (Fig. 8b and c). In the visualizations, he recognized that miR-200c seemed to play an important role since many mRNAs that were overexpressed in invasive breast cancer cell lines were targeted by the miR-200c that was significantly under-regulated in invasive breast cancer cell lines. After manually reviewing previous literature of each miRNA, he confirmed that most of the miRNAs including miR-200c were known to affect invasion in breast cancer cells (i.e., miR-34a, miR-141, miR-183, miR-200c, and miR-203 [23–26]). Then, he was interested in significantly under-regulated target mRNAs. In the treemap visualization, he could find 49 mRNAs including known target genes that were linked to cancer cell invasion ability (i.e., Cfl2 [25] and PRKCA [27]). He said that any other novel target mRNAs including FAM55C and SCHIP1 would also play an important role in breast cancer, and it would be valuable to further analyze them (e.g., Tissue microarray analysis).

During the post-study interview, the participant gave us positive comments on miRTarVis+. First, he said the main advantage of miRTarVis+ is its support of iterative analysis: Users can flexibly go back to previous steps (e.g., *filter* or *predict* steps) in the analysis pipeline, change some settings, and visualize the resulting miRNA-mRNA interactions again. He said that in other tools (e.g., magia2 [6]), he had to go all the way back to the *data load* step when he simply wants to change filtering or prediction parameters. Second, he positively commented on the treemap visualization. He said that although he had no knowledge about treemap before the study, he could learn about the visualization with only a short tutorial session (i.e., 15 min). Moreover, he said that clearly visualizing fold change values without occlusion helped him with more efficient analysis.

4. Conclusions

We introduced miRTarVis+, which is an enhanced version of our previous tool, miRTarVis [8]. miRTarVis+ is a Web-based interactive visual analytics tool for miRNA-mRNA expression profile data. We defined a representative data analysis pipeline and designed miRTarVis+ to support the analysis pipeline using a step-by-step navigation interface. miRTarVis+ integrates various miRNA target prediction algorithms, including a novel one using the MINE analysis, into the analysis pipeline. We provided an overview of multiple prediction results where users can effectively select miRNA-mRNA pairs of their interest. miRTarVis+ shows the selected miRNA target network using an interactive treemap and node-link diagram.

We conducted two case studies to prove the effectiveness of miRTarVis+. In the first study, we analyzed a miRNA-mRNA expression profile data of asthma patients and found a potentially novel mechanism by which adiposity increases fibrosis in asthma. Through the second case study, we also showed that our system enabled users to find miRNA-mRNA interactions which may play important role in breast cancer dataset.

In the future work, we will improve miRTarVis+ by integrating transcription factors (TFs), which regulate gene expression with miRNAs, into the analysis procedure of miRTarVis+. We will conduct further study on visualization techniques for the gene expression regulatory network and integrate new visualization techniques into miRTarVis+. We will also research on a new interaction technique to solve a problem of the Treemap, namely, the lack of the visual cue to show common targets of a miRNA. In addition, we will conduct a controlled user study to miRNA researchers to verify the effectiveness of our interactive visualizations by comparing them with other visualizations in existing tools (e.g., node-link diagrams in Magia2 [6]).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (No. NRF-2014R1A2A2A03006998 and NRF-2016R1A2B2007153) and by the Hankuk University of Foreign Studies Research Fund. The ICT at Seoul National University provided research facilities for this study.

References

[1] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2) (2004) 281–297.
[2] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (1) (2005) 15–20.
[3] A.J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, D.S. Marks, MicroRNA targets in Drosophila, Genome Biol. 5 (1) (2004) R1.
[4] S. Nam, M. Li, K. Choi, C. Balch, S. Kim, N.P. Nephew, MicroRNA and mRNA integrated analysis (MMIA): A web tool for examining biological functions of microRNA expression, Nucleic Acids Res. 37 (Suppl. 2) (2009) W356–W362.
[5] G.T. Huang, C. Athanassiou, P.V. Benos, mirConnX: condition-specific mRNA-microRNA network integrator, Nucleic Acids Res. 39 (Suppl. 2) (2011) W416–W423.
[6] A. Biosognin, G. Sales, A. Coppe, S. Bortoluzzi, C. Romualdi, MAGIA2: from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits, Nucleic Acids Res. 40 (W1) (2012) W13–W21.
[7] J. Huang, F. Gutierrez, H.J. Strachan, D. Dou, W. Huang, B. Smith, J.A. Blake, K. Eilbeck, D.A. Natale, Y. Lin, B. Wu, N. de Silva, X. Wang, Z. Liu, G.M. Borchert, M. Tan, A. Ruttenberg, OmniSearch: A semantic search system based on the Ontology for MIcroRNA Target (OMIT) for microRNA-target gene interaction data, J. Biomed. Semantics 7 (1) (2016) 25.
[8] D. Jung, B. Kim, R.J. Freishtat, M. Giri, E. Hoffman, J. Seo, miRTarVis: an interactive visual analysis tool for microRNA-mRNA expression profile data, BMC Proc. 9 (6) (2015) S2.
[9] J.C. Huang, T. Babak, T.W. Corson, G. Chua, S. Khan, B.L. Gallie, et al., Using expression profiling data to identify human microRNA targets, Nat. Method 4 (12) (2007) 1045–1049.
[10] D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, et al., Detecting novel association in large data sets, Science 334 (6062) (2011) 1518–1524.
[11] A. Kozomara, S. Griffiths-Jones, miRBase: Annotating high confidence microRNAs using deep sequencing data, Nucleic Acids Res. 42 (D1) (2013) D68–D73.
[12] Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, et al., miR2Disease: a manually curated database for microRNA deregulation in human disease, Nucleic Acids Res. 37 (Suppl. 1) (2009) D98–D104.
[13] M. Safran, I. Dalah, J. Alexander, N. Rosen, T.I. Stein, M. Shmoish et al., GeneCards version 3: the human gene integrator, Database baq020 (2010).
[14] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, H. Pfister, UpSet: visualization of intersecting sets, IEEE Trans. Visualization Comput. Graph. 20 (12) (2014) 1983–1992.
[15] B. Kim, B. Lee, J. Seo, Visualizing set concordance with permutation matrices and fan diagrams, Interact. Comput. 19 (5–6) (2007) 630–643.
[16] B. Kim, B. Lee, S. Knoblach, E. Hoffman, J. Seo. GeneShelf: A Web-based visual interface for large gene expression time-series data repositories. IEEE Trans. Visualization Comput. Graph. 15(6) (2009).
[17] D3-Labeler. Retrieved April 1. 2017 from https://github.com/tinker10/D3-Labeler/.
[18] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E 69 (6) (2004) 066138.
[19] D. Betal, M. Wilson, A. Gabow, D.S. Marks, C. Sander, The microRNA.org resource: targets and expression, Nucleic Acids Res. 36 (Suppl. 1) (2008) D149–D153.

[20] S.C. Ferrante, E.P. Nadler, D.K. Pillai, M.J. Hubal, Z. Wang, J.M. Wang et al., Adipocyte-derived exosomal miRNAs: a novel mechanism for obesity-related disease, Pediatr. Res. (2015) 10.1038, pr.2014.202.
[21] N.B. Ben-Moshe, R. Avraham, M. Kedmi, A. Zeisel, A. Yitzhaky, Y. Yarden, E. Domany, Context-specific microRNA analysis: identification of functional microRNAs and their mRNA targets, Nucleic Acids Res. 40 (21) (2012) 10614–10627.
[22] D. Luo, J.M. Wilson, N. Harvel, J. Liu, L. Pei, S. Huang, L.A. Hawthorn, H. Shi, A systematic evaluation of miRNA: mRNA interactions involved in the migration and invasion of breast cancer cells, J. Transl. Med. 11 (1) (2013) 57.
[23] L. Li, Y. Linjin, L. Jinmei, G. Jie, G. Jiaoli, X. Xiaoming, MiR-34a inhibits proliferation and migration of breast cancer through down-regulation of Bcl-2 and SIRT1, Clin. Exp. Med. 13 (2) (2013) 109–117.
[24] P.A. Gregory, A.G. Bert, E.L. Paterson, S.C. Barry, A. Tsykin, G. Farshid, M.A. Vadas, Y. Khew-Goodall, G.J. Goodall, The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1, Nat. Cell Biol. 10 (2008) 593–601.
[25] D. Luo, M.W. James, H. Nikki, L. Jimei, P. Lirong, H. Shuang, H. LesleyAnn, S. Huidong, A systematic evaluation of miRNA: mRNA interactions involved in the migration and invasion of breast cancer cells, J. Transl. Med. 11 (1) (2013) 57.
[26] P. Li, S. Cheng, H. Lingling, Z. Hui, H. Lihua, C. Zeneng, Z. Qubo, MiR-183/-96/-182 cluster is up-regulated in most breast cancers and increases cell proliferation and migration, Breast Cancer Res. 16 (6) (2014) 473.
[27] G. Lonne, L. Cornmark, I. Zahirovic, G. Landberg, K. Jirstrom, C. Larsson, PKCα expression is a marker for breast cancer aggressiveness, Mol. Cancer 9 (1) (2010) 76.