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