

HisCoM-mimi: software for hierarchical structural component analysis for miRNA-mRNA integration model for binary phenotypes

Yongkang Kim¹, Taesung Park^{1,2*}

¹Department of Statistics, Seoul National University, Seoul 08826, Korea

²Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea

To identify miRNA-mRNA interaction pairs associated with binary phenotypes, we propose a hierarchical structural component model for miRNA-mRNA integration (HisCoM-mimi). Information on known mRNA targets provided by TargetScan is used to perform HisCoM-mimi. However, multiple databases can be used to find miRNA-mRNA signatures with known biological information through different algorithms. To take these additional databases into account, we present our advanced application software for HisCoM-mimi for binary phenotypes. The proposed HisCoM-mimi supports both TargetScan and miRTarBase, which provides manually-verified information initially gathered by text-mining the literature. By integrating information from miRTarBase into HisCoM-mimi, a broad range of target information derived from the research literature can be analyzed. Another improvement of the new HisCoM-mimi approach is the inclusion of updated algorithms to provide the lasso and elastic-net penalties for users who want to fit a model with a smaller number of selected miRNAs and mRNAs. We expect that our HisCoM-mimi software will make advanced methods accessible to researchers who want to identify miRNA-mRNA interaction pairs related with binary phenotypes.

Keywords: integration analysis, miRNA, miRNA database, mRNA

Availability: HisCoM-mimi is available at <http://statgen.snu.ac.kr/software/hiscom-mimi/>.

Introduction

miRNA is a well-known form of noncoding RNA that affects biological mechanisms by regulating the expression of target mRNA. Many researchers have found that cancer cells and normal cells exhibit different inhibition mechanisms, suggesting that miRNAs could be used as biological markers for the diagnosis of cancer [1-3]. In our previous study, we presented a hierarchical structural component model (HisCoM-mimi) to find the miRNA-mRNA interaction pairs associated with binary phenotypes that could be candidates for cancer diagnosis biomarkers with an interpretable biological inhibition mechanism. Recently, many findings regarding the target mRNAs of miRNAs have been incorporated into various databases. TargetScan is one such database with recently updated findings [4]. The basic principle used by TargetScan to predict the mRNAs that miRNAs target for inhibition is to compare the sequences of untranslated mRNA regions to those of miRNAs [5]. However, many studies have shown that miRNAs select their target mRNAs based not only on the similarity of sequences, but also on other structural findings [6]. Thus, miRNA-mRNA integration analysis requires experimental confirmation of which mRNAs are truly inactivated by miRNAs. MiRTarBase is a data-

base that collects experimental findings [6]. In our previous study, we only used TargetScan to find pairs of miRNA-mRNA relationships [7]. To enable researchers to utilize more flexibly information regarding the target mRNAs inhibited by miRNAs, we added miRTarBase database information to our software.

Implementation

Fig. 1 shows the hierarchical structural component analysis work-

flow for the HisCoM-mimi application, which requires miRNA and mRNA expression datasets and additional files (phenotype and covariates). The program now accepts two formats (miRNA and mRNA CEL files or an Excel-type expression dataset).

Next, miRNA-mRNA networks are constructed by combining the miRNA database information and correlation coefficients computed based on the user-entered datasets. Users can select an miRNA database in three ways: TargetScan results, miRTarBase results, and the intersection of both databases. The user can define the fil-

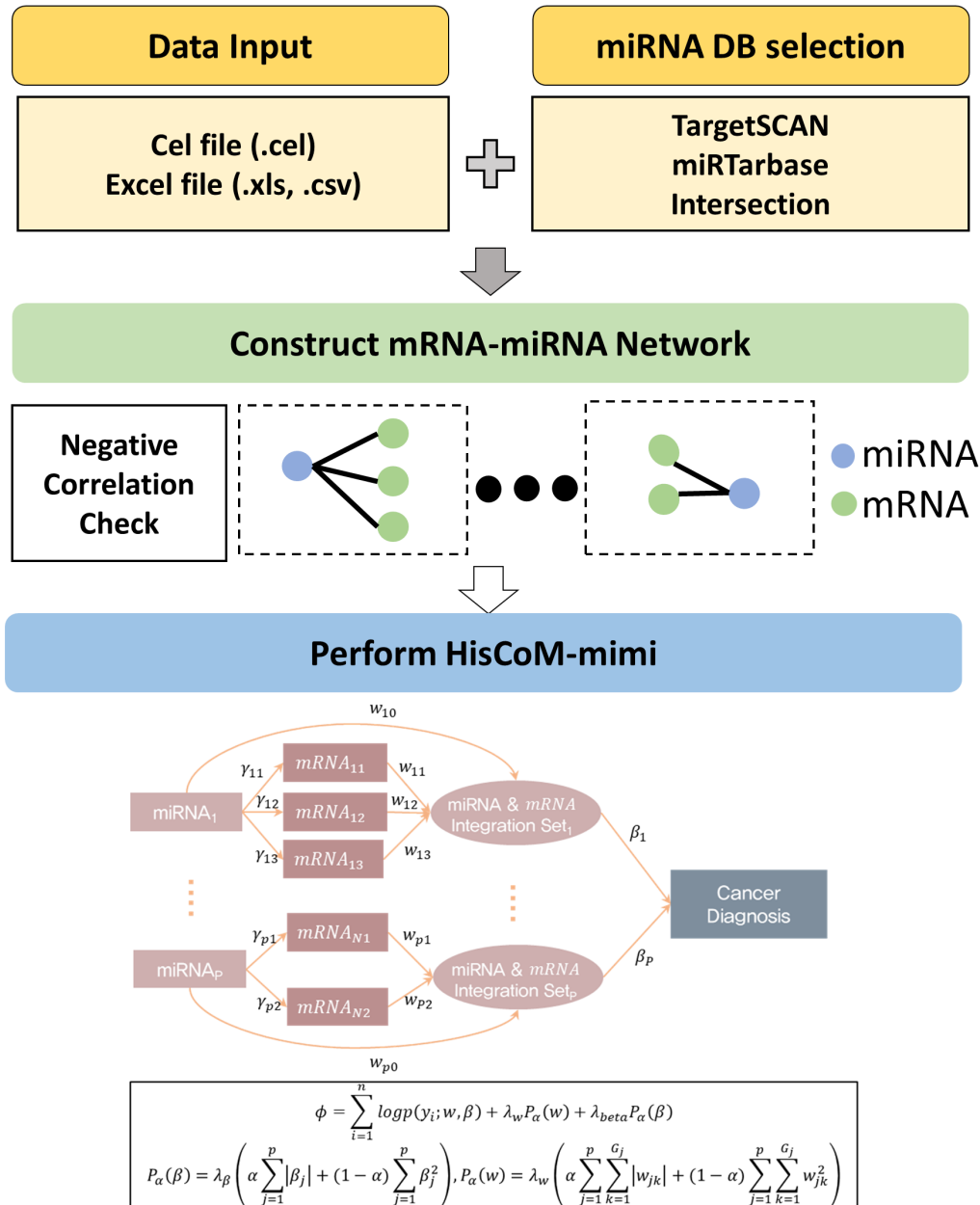


Fig. 1. Workflow of HisCoM-mimi. The hierarchical structural component analysis workflow for the HisCoM-mimi application is shown. HisCoM-mimi first constructs an mRNA-miRNA network that connects negatively correlated miRNAs and target mRNAs. It then estimates the model parameters by maximizing the objective function via the alternating least square method with double-penalty functions for the coefficients. HisCoM-mimi, hierarchical structural component model for miRNA-mRNA integration.

tering network criteria by two options: (1) the choice of the databases and (2) a p-value threshold for the correlation coefficients between miRNAs and mRNAs.

After constructing an mRNA-miRNA integration set, HisCoM-mimi can be performed. HisCoM-mimi can accommodate not only the ridge penalty, but also the lasso and elastic-net penalties, which result in reduced computing time and a smaller number of nonzero coefficients [8,9]. A cross-validation procedure is necessary to find the optimal penalties that maximize the log likelihood of the validation set [7].

Conclusion

In this paper, we introduced our HisCoM-mimi software for miRNA-mRNA integration analysis. The current HisCoM-mimi application can use both the TargetScan and miRTarBase databases. Furthermore, in addition to the ridge penalty, HisCoM-mimi can accommodate lasso and elastic-net penalties.

ORCID

Yongkang Kim: <https://orcid.org/0000-0002-8837-2620>

Taesung Park: <https://orcid.org/0000-0002-8294-590X>

Authors' Contribution

Conceptualization: TP. Data curation: YK. Funding acquisition: TP. Methodology: YK, TP. Writing - original draft: YK. Writing - review & editing: TP.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI16C2037) and the Bio-Synergy Research Project of the Ministry of Science, ICT and Future Planning through the National Research Foundation (grant number: 2013M3A9C4078158)

References

1. Cho JH, Gelinias R, Wang K, Etheridge A, Piper MG, Batte K, et al. Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. *BMC Med Genomics* 2011;4:8.
2. Farazi TA, Hoell JI, Morozov P, Tuschl T. MicroRNAs in human cancer. *Adv Exp Med Biol* 2013;774:1–20.
3. Kang SM, Lee HJ. MicroRNAs in human lung cancer. *Exp Biol Med (Maywood)* 2014;239:1505–1513.
4. Shi Y, Yang F, Wei S, Xu G. Identification of key genes affecting results of hyperthermia in osteosarcoma based on integrative ChIP-Seq/TargetScan analysis. *Med Sci Monit* 2017;23:2042–2048.
5. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120:15–20.
6. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* 2011;39:D163–D169.
7. Kim Y, Lee S, Choi S, Jang JY, Park T. Hierarchical structural component modeling of microRNA-mRNA integration analysis. *BMC Bioinformatics* 2018;19:75.
8. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Methodol* 1996;58:267–288.
9. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B* 2005;67:301–320.