Faculty of Engineering and Information Technology
University of Technology, Sydney

Rule Mining on MicroRNA Expression Profiles for Human Disease Understanding

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Renhua Song

July 2016

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgments

First, and foremost, I would like to express my gratitude to my chief supervisor, Assoc. Prof. Jinyan Li, and to my co-supervisors Assoc. Prof. Paul Kennedy and Assoc. Prof. Daniel Catchpoole. I am extremely grateful for all the advice and guidance so unselfishly given to me over the last three and half years by these three distinguished academics. This research would not have been possible without their high order supervision, support, assistance and leadership.

The wonderful support and assistance provided by many people during this research is very much appreciated by me and my family. I am very grateful for the help that I received from all of these people but there are some special individuals that I must thank by name.

A very sincere thank you is definitely owed to my loving husband Jing Liu, not only for his tremendous support during the last three and half years, but also his unfailing and often sorely tested patience. My deepest thanks is also extended to every member of my family, especially my little son Boao Liu, for his love, understanding and obedience. Special thanks are also owed to my parents Congchun Song and Guiyun Zhang for all their tremendous assistance and unfailing support.

My sincere appreciation and gratitude goes to Dr. Qian Liu and Dr. Yun Zheng, for all their freely given invaluable advice and insightful discussion throughout my research. My heartfelt admiration and thanks is also extended to Dr. Gyorgy Hutvagner and Dr. Hung Nguyen in the Centre for Health Technologies of UTS, Dr. Kotagiri Ramamohanarao in the University of

Melbourne, Dr. Limsoon Wong in the National University of Singapore, Dr. Tao Liu in the Children's Cancer Institute Australia. Their respective assistance in terms of biological knowledge was invaluable.

I would like to thank of my colleagues at the Advanced Analytics Institute (AAI), for their selfless support over the course of my PhD candidature. and for all the fun that we shared in the last three and half years.

Thanks and praises are also due to Almighty God, my Lord and Saviour, who has bestowed great blessing upon me throughout my life. During the last three and half years, while conducting my research and preparing to write this dissertation, I have come to realise the great relevance of the words of the apostle Saint Paul who proclaimed, "I can do all things through Christ who strengthens me".

Finally, to anyone whom I have not mentioned, please forgive me. I can most definitely assure you that you have occupied a unique and special place in my thoughts. Thank you!

Renhua Song
July 2016 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Below is the list of journal and conference papers associated with my PhD research:

Journal Papers Published

- Zheng, Y., Ji, B., **Song, R.**, Wang, S., Li, T., Zhang, X., Chen, K., Li, T. and Li, J., 2016. Accurate detection for a wide range of mutation and editing sites of microRNAs from small RNA high-throughput sequencing profiles. Nucleic acids research, p.gkw471. `http://dx.doi.org/10.1093/nar/gkw471`

- Tee, A.E., Liu, B., **Song, R.**, Li, J., Pasquier, E., Cheung, B.B., Jiang, C., Marshall, G.M., Haber, M., Norris, M.D. and Fletcher, J.I., 2016. The long noncoding RNA MALAT1 promotes tumor-driven angiogenesis by up-regulating pro-angiogenic gene expression. Oncotarget, `http://dx.doi.org/10.18632/oncotarget.6675`

- Tee, A.E., Liu, P., Maag, J., **Song, R.**, Li, J., Cheung, B.B., Haber, M., Norris, M.D., Marshall, G.M., Dinger, M. and Liu, T., 2015. The long noncoding RNA MALAT1 promotes hypoxia-driven angiogenesis by upregulating pro-angiogenic gene expression in neuroblastoma cells. Cancer Research, 75(Supplement 15), pp.146-146

- Liu, Q., **Song, R.** and Li, J., 2015. Inference of gene interaction networks using conserved subsequential patterns from multiple time

course gene expression datasets. BMC Genomics, 16(Supplement 12): S4

- **Song, R.**, Catchpoole, D.R., Kennedy, P.J. and Li, J., 2015. Identification of lung cancer miRNAmiRNA co-regulation networks through a progressive data refining approach. Journal of Theoretical Biology, 380, pp.271-279

- **Song, R.**, Liu, Q., Liu, T. and Li, J., 2015. Connecting rules from paired miRNA and mRNA expression data sets of HCV patients to detect both inverse and positive regulatory relationships. BMC Genomics, 16(Supplement 2): S11

- **Song, R.**, Liu, Q., Hutvagner, G., Nguyen, H., Ramamohanarao, K., Wong, L. and Li, J., 2014. Rule discovery and distance separation to detect reliable miRNA biomarkers for the diagnosis of lung squamous cell carcinoma. BMC Genomics, 15(Supplement 9): S16

Conference Papers Presented

- Liu, Q., **Song, R.** and Li, J. "Inference of gene interaction networks using conserved subsequential patterns from multiple time course gene expression datasets." `www.jsbi.org/giw-incob2015/`, Tokyo, Japan. **Oral Presentation**

- **Song, R.**, Liu, Q., Liu, T. and Li, J. "Connecting rules from paired miRNA and mRNA expression data sets of HCV patients to detect both inverse and positive regulatory relationships." `http://apbc2015.mbc.nctu.edu.tw/`, HsinChu, Taiwan. **Oral Presentation.**

- **Song, R.**, Liu, Q., Hutvagner, G., Nguyen, H., Ramamohanarao, K., Wong, L. and Li, J. "Rule discovery and distance separation to detect reliable miRNA biomarkers for the diagnosis of lung squamous cell carcinoma."`http://incob.apbionet.org/incob14/`, Sydney, Australia. **Oral and Poster Presentation.**

Papers to be Submitted/Under Review

- **Song R.**, Zheng, Y., Catchpoole, D.R., Kennedy, P.J. and Li, J. "A Novel Framework for Inferring Self-regulation miRNAs." **Under Revision**

Abstract

This research employs rule mining methods to study the important roles of miRNAs in human diseases. From past experience and from reviewing the literature, rule mining is a widely used data mining technique for the discovery of interesting relationships in large data sets. MicroRNAs (miRNAs) are endogenous and highly conserved non-coding RNA molecules. They can inhibit and/or promote the post-transcriptional expression of target messenger RNAs (mRNAs). miRNAs thus play a pivotal role in a cell's differentiation, proliferation, growth, mobility, and apoptosis, as well as in viral replication and proliferation. This has inspired many research works aimed at detecting miRNAs' functions in human disease. However, with the current deluge of miRNA data, previous works have suffered from limitations in terms of handling the relationship between various molecules. Firstly, they usually identify single miRNAs as biomarkers, and always produce low sensitivity and specificity. Secondly, intensive research largely depends on the inverse expression relationships between miRNAs and mRNAs to discover miRNA-mRNA regulatory modules. Finally, the miRNA-miRNA co-regulations and miRNA self-regulations have not been well investigated. As a result, rule mining is a powerful new technology with great potential to help researchers focus on the most important miRNAs for understanding human diseases. This thesis reports our past and current research outcomes in this area. The contributions of the thesis are as follows:

- A novel rule mining method is proposed to detect the significant miRNA biomarkers.

- A "change to change" method is proposed to mine both positive and negative regulatory relationships from paired miRNA and mRNA expression data sets.
- A progressive data refining approach is proposed to identify the lung cancer miRNA-miRNA co-regulation network.
- A novel framework is proposed to detect the self-regulation miRNAs.

The research was conducted through four case studies. (1) The first case study was on lung squamous cell carcinoma for accurate diagnosis of this disease through the reliable miRNA biomarkers identified by a novel rule discovery method. (2) The second case study was on paired miRNA and mRNA expression data of HCV patients to detect both positive and negative regulatory modules. (3) The third case study was on lung cancer data sets for the computational methods to identify miRNA-miRNA co-regulation networks and miRNA-miRNA co-regulatory relationships. (4) The fourth case study was on multiple data types to infer self-regulation miRNAs in humans through an integrative rule mining framework and approach. All the results have been verified by the existing literature and databases.