An Integrative Analysis of ncRNA-mRNA Using Co-expression Network to Discover Potential Contributions of Coding-non-coding RNA Clusters[1]

Li Guo, Yang Zhao, Sheng Yang, Hui Zhang, Feng Chen

Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China

LG: lguo@njmu.edu.cn, FC: fengchen@njmu.edu.cn

Abstract. Non-coding RNAs (ncRNAs), especially for microRNAs (miRNAs), have been widely studied as crucial negative regulatory molecules. Long non-coding RNAs (lncRNAs) also have attracted the attention of researchers due to their potential contribution in multiple essential biological processes. To understand the potential interactions between miRNAs, lncRNAs and mRNAs and their potential roles in tumorigenesis, we reported an integrative analysis to predict clustered ncRNA-mRNA with consistent functions, and predict clusters at the single molecule level. The method aims to discover those potential clusters of coding-non-coding RNAs that maybe contribute to occurrence and development of human diseases. Based on expression profiles and abnormal expression profiles of miRNAs, lncRNAs and mRNAs, co-expression network analysis can be performed at the single molecule and multiple RNA molecules, respectively. Some clustered RNAs at the single RNA molecule can be obtained, and these members always have consistent functions. Although these non-coding RNAs or coding RNAs are analyzed at the single molecule level, they have close functional relationships, especially between miRNAs and their target mRNAs. Therefore, based on their potential functional and sequence relationships, further coding-non-coding co-expression network can be constructed based on integrative expression and functional analysis across different molecule levels. The comparison analysis of the single and multiple molecules will provide more information to predict interaction between miRNAs and lncRNAs, ncRNAs and mRNAs. Furthermore, based on special miRNA group

[1]This work was supported by the National Natural Science Foundation of China (No. 61301251, 81072389 and 81373102), the Research Fund for the Doctoral Program of Higher Education of China (No. 211323411002 and 20133234120009), the China Postdoctoral Science Foundation funded project (No. 2012M521100), the key Grant of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 10KJA33034), the National Natural Science Foundation of Jiangsu (No. BK20130885), the Natural Science Foundation of the Jiangsu Higher Education Institutions (No. 12KJB310003 and 13KJB330003), the Jiangsu Planned Projects for Postdoctoral Research Funds (No. 1201022B), the Science and Technology Development Fund Key Project of Nanjing Medical University (No. 2012NJMU001), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

with potential functional relationships, such as miRNA gene cluster and gene family, especially for complex miRNA processing and maturation mechanism, and potential miRNA-miRNA interaction, miRNAs are involved in the more complex regulatory network and pattern. The location distributions and potential sequence correlations are also analyzed across different RNA molecules, which maybe implicate the potential functional relationships between ncRNAs and mRNAs, especially between miRNAs and lncRNAs. The systematic and large-scale co-expression networks based on the single and multiple RNA molecule levels will provide more association of coding-non-coding RNAs and their potential contributions in occurrence and development of diseases. Finally, the analysis also implicates the functional and evolutionary relationships across coding and non-coding RNAs.

Keywords: miRNA; lncRNA; mRNA; integrative analysis; co-expression

1 Introduction

Non-coding RNAs (ncRNAs) have been widely studied because their versatile and essential biological roles. Of these, microRNAs (miRNAs) are a class of major negative regulators at the post-transcriptional level [1]. They are shorter (~22-nt) ncRNAs and phylogenetically well-conserved in different animal species. They can regulate gene expression through binding to complementary sequences in their target mRNAs, which promotes translational repression or mRNA degradation [2, 3]. The small ncRNAs show crucial biological roles as key negative regulators of gene expression, and they contribute to many essential biological processes [4] and occurrence and development of many human diseases [5]. Similarly, long non-coding RNAs (lncRNAs) also have attracted many researches because their potential biological roles, such as they may be potential regulatory molecule as well as miRNAs [6, 7]. These ncRNAs have close correlations with coding RNAs, mRNAs.

Generally, we perform expression analysis of the single molecule level (such as mRNA, miRNA, lncRNA, and etc.) to discover potential mechanism in the occurrence and development of human diseases, despite any disease contains complex and abnormal biological processes that are involved in multiple molecules, including many ncRNAs and coding RNAs and potential interactions across different molecules. Therefore, it is not enough to understand the complex mechanism via simple analysis of the single RNA molecule level. The systematic and genome-wide analysis based on co-expression network analysis across multiple RNA molecules levels is quite necessary to unveil potential mechanisms in human diseases.

By applying gene co-expression analysis, we can predict clustered genes that may be relevant to cancer development and prognosis. Members in a cluster always have consistent functions and biological roles. Potential biomarker in different types of cancers can be predicted by using co-expression analysis. As crucial regulatory molecules, such as miRNA and lncRNA, they play key roles in gene expression that contribute to multiple essential biological processes. Therefore, systematic analysis across different molecule levels is quite essential to unveil potential complex mechanism in tumorigenesis. Herein, we reported a method to predict clusters of mRNAs,

lncRNAs and miRNAs in the single and multiple molecule level. To further understand potential interaction between different molecules, we also attempted to obtain clustered coding-non-coding RNAs via an integrative analysis across different RNA molecules. Furthermore, based on miRNA gene clusters and gene families, a series of miRNAs can be analyzed through their potential close functional and physical relationships as well as miRNA processing and maturation mechanisms.

2 Materials and Methods

According to obtained abundantly expressed miRNA, lncRNA and mRNA datasets, an integrative analysis can be performed based on their potential expression relationships using correlation analysis, simultaneously including their functional relationships and location distributions on chromosomes (Fig. 1). Co-expression network analysis will be constructed at the single molecule and multiple RNA molecules. Some special RNA groups should be considered according to their potential functional relationships. For example, miRNA gene clusters and gene families (Fig. 1). These special miRNAs maybe co-regulate or co-contribute to multiple biological processes, and they are thought with the same ancestor miRNA genes. Although the co-expression network based on the single molecule will be focused on mRNA level, the theoretical and virtual analysis may be consistent or inconsistent, especially they always are involved in complex regulatory patterns. Co-expression network of multiple RNA molecules, coding-non-coding RNAs based on integrative analysis, will show a complex regulatory network of miRNA, lncRNA and mRNA expression profiles. The comparison of results of the single and multiple RNAs will indicate the potential relationships across coding and non-coding, and different non-coding RNAs (Fig. 1).

3 Results and Discussion

Next-generation sequencing techniques provide an opportunity to obtain genome-wide RNA expression profiles, especially for those non-coding RNAs. As crucial regulatory molecules, these ncRNAs play important roles in multiple biological processes via targeting messenger RNAs (mRNAs). The analysis based on the single RNA molecule has been widely studied, but it is not enough to unveil the complex relationships between different RNAs, particular between ncRNAs and mRNAs. Indeed, functions of some ncRNAs, especially for lncRNAs, are still remain largely unknown. Any abnormal biological pathway may be involved in a series of aberrantly expressed ncRNAs. The regulatory patterns lead to more complex regulatory network in vivo across different RNA molecules. Therefore, it is necessary to unveil their potential relationships, especially for predict functional clusters that may be relevant to cancer development and prognosis.

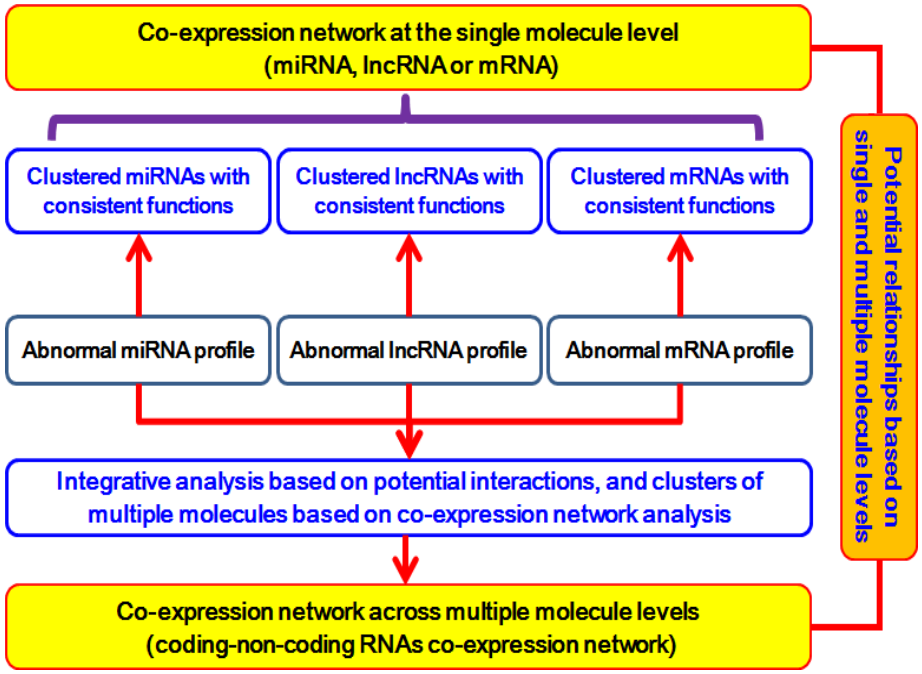

Fig. 1. A flowchart of co-expression analysis for the single molecule and integrated analysis of multiple molecules.

According to miRNA, lncRNA and mRNA expression profiles, co-expression network analysis can be used to obtain potential functional clusters and unveil the potential functional relationships across multiple RNAs. Due to the complexity of regulatory network, the analysis can be performed based on the single and multiple RNA molecules. Firstly, each RNA molecule is constructed co-expression network based on abundantly expression profiles and potential relationships with mRNAs (Fig. 1). For example, miRNAs play biological roles via negatively regulating their target mRNAs. As a class of small and flexible non-coding RNAs, miRNAs always co-regulate the same pathway by forming some miRNA groups (such as miRNA gene clusters and gene families) [8]. These clustered and/or homologous miRNAs may be derived from complex historic duplication processes, and restrict the evolutionary divergence with the similar or same functional regions (nucleotides 2-8, seed sequences). Moreover, multiple isomiRs can be yielded from the miRNA locus due to alternative and imprecise cleavage of Drosha and Dicer [9, 10]. miRNA-miRNA interaction maybe contribute to dynamic miRNA expression profiles and regulatory patterns [11, 12]. All of these characteristics should be considered, although the theoretical and virtual expression patterns of mRNAs are always inconsistent due to involved in complex regulatory networks. Secondly, an integrative analysis across different RNA molecules is performed based on their potential relationships, including the functional, sequence and location distribution relationships. The sequence similarity always implicates the

potential binding event and regulatory patterns between ncRNAs and mRNAs. The further co-expression network of coding-non-coding RNAs can be constructed based on the integrative analysis. The clustered RNAs imply the potential relationships between different RNA molecules, although ncRNAs always negatively regulate their target mRNAs. Finally, the networks of the single RNA molecule and multiple RNA molecules will indicate the potential correlations by comparison analysis.

Furthermore, we can construct the co-expression network by using the abnormal RNA expression profiles in the diseased samples or treated samples. The aberrantly expressed ncRNA and mRNA profiles will provide more direct information in the occurrence and development of diseases. NcRNAs can negatively regulate coding RNAs, although different ncRNAs also maybe have close functional relationships. Each miRNA may have thousands of target mRNAs, and each mRNA also may be regulated by multiple miRNAs, such as miRNA gene clusters and gene families can co-regulate biological process. Different miRNAs also show potential interactions based on reverse complementarily binding events between miRNAs, especially between sense and antisense miRNAs [11, 12]. miRNAs and their target mRNAs show flexible expression, although they should show opposite expression patterns via miRNA-mRNA interaction. Therefore, the comprehensive analysis based on the aberrantly miRNA, lncRNA and mRNA profiles will provide the actual abnormal expression in the diseased or treated samples. Moreover, different RNA molecules show potential evolutionary relationships. The clustered co-expression network also implicates their phylogenetic relationships, which also provide more information to unveil the potential functional implications. Most of lncRNAs still remain largely unknown, while the co-expression network analysis of multiple RNA molecules can provide the implication of lncRNAs. Taken together, the co-expression network analysis based on the integrative analysis of multiple RNAs, coding and non-coding RNAs, will provide more information about unveiling potential coding-non-coding RNA clusters and functional and evolutionary implication.

References

1. D. P. Bartel, "MicroRNAs: Genomics, biogenesis, mechanism, and function," *Cell,* vol. 116, pp. 281-297, Jan 23 2004.
2. R. W. Carthew and E. J. Sontheimer, "Origins and Mechanisms of miRNAs and siRNAs," *Cell,* vol. 136, pp. 642-655, Feb 20 2009.
3. H. Guo, N. T. Ingolia, J. S. Weissman, and D. P. Bartel, "Mammalian microRNAs predominantly act to decrease target mRNA levels," *Nature,* vol. 466, pp. 835-40, Aug 12 2010.
4. G. Stefani and F. J. Slack, "Small non-coding RNAs in animal development," *Nature Reviews Molecular Cell Biology,* vol. 9, pp. 219-230, Mar 2008.
5. C. Xiao and K. Rajewsky, "MicroRNA control in lymphocyte physiology and pathology," *Annals of Oncology,* vol. 19, pp. 117-117, Jun 2008.
6. K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Mol Cell,* vol. 43, pp. 904-14, Sep 16 2011.

7. M. Huarte, M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks, and J. L. Rinn, "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response," *Cell,* vol. 142, pp. 409-19, Aug 6 2010.
8. S. R. Viswanathan, C. H. Mermel, J. Lu, C. W. Lu, T. R. Golub, and G. Q. Daley, "microRNA Expression during Trophectoderm Specification," *Plos One,* vol. 4, p. e6143, Jul 3 2009.
9. R. D. Morin, M. D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A. L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. J. Eaves, and M. A. Marra, "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells," *Genome Research,* vol. 18, pp. 610-621, Apr 2008.
10. J. G. Ruby, C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel, "Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C-elegans," *Cell,* vol. 127, pp. 1193-1207, Dec 15 2006.
11. L. Guo, B. Sun, Q. Wu, S. Yang, and F. Chen, "miRNA-miRNA interaction implicates for potential mutual regulatory pattern," *Gene,* vol. 511, pp. 187-94, Dec 15 2012.
12. L. Guo, T. Liang, W. Gu, Y. Xu, Y. Bai, and Z. Lu, "Cross-mapping events in miRNAs reveal potential miRNA-mimics and evolutionary implications," *PLoS ONE,* vol. 6, p. e20517, 2011.