

Research Article

Global Analysis of miRNA Gene Clusters and Gene Families Reveals Dynamic and Coordinated Expression

Li Guo,¹ Sheng Yang,¹ Yang Zhao,¹ Hui Zhang,¹ Qian Wu,² and Feng Chen¹

¹ Department of Epidemiology and Biostatistics and Ministry of Education Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing 211166, China

² State Key Laboratory of Reproductive Medicine, Department of Hygienic Analysis and Detection, School of Public Health, Nanjing Medical University, Nanjing 211166, China

Correspondence should be addressed to Qian Wu; wuqian@njmu.edu.cn and Feng Chen; fengchen@njmu.edu.cn

Received 17 January 2014; Accepted 26 February 2014; Published 25 March 2014

Academic Editor: Xing-Ming Zhao

Copyright © 2014 Li Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To further understand the potential expression relationships of miRNAs in miRNA gene clusters and gene families, a global analysis was performed in 4 paired tumor (breast cancer) and adjacent normal tissue samples using deep sequencing datasets. The compositions of miRNA gene clusters and families are not random, and clustered and homologous miRNAs may have close relationships with overlapped miRNA species. Members in the miRNA group always had various expression levels, and even some showed larger expression divergence. Despite the dynamic expression as well as individual difference, these miRNAs always indicated consistent or similar deregulation patterns. The consistent deregulation expression may contribute to dynamic and coordinated interaction between different miRNAs in regulatory network. Further, we found that those clustered or homologous miRNAs that were also identified as sense and antisense miRNAs showed larger expression divergence. miRNA gene clusters and families indicated important biological roles, and the specific distribution and expression further enrich and ensure the flexible and robust regulatory network.

1. Introduction

The small non-coding RNA regulatory molecules, microRNAs (miRNAs), play an important role in multiple biological processes through negatively regulating gene expression [1]. Abnormally expressed miRNAs may contribute to various human diseases, including cancer development, and some have been identified as potential oncomiRs or tumor suppressors [2, 3]. Some miRNAs are preferentially located at fragile sites and regions and are abnormally expressed in cancer samples [4]. Those deregulated miRNAs have been widely studied as potential biomarkers, especially for circulating miRNAs in human diseases [5–7].

miRNAs in gene cluster or family may have functional relationships via coregulating or coordinately regulating biological processes [8, 9], although they have various expression levels due to complex maturation and degradation mechanisms [10–12]. These clustered miRNAs are quite popular in

metazoan genomes, and they may be involved in homologous miRNA genes via duplication evolutionary histories [13–15]. Simultaneously, the phenomenon of multicopy miRNA precursors (pre-miRNAs) further complicates the distributions of miRNA gene cluster and family and also implicates the dynamic evolutionary process in the miRNA world [15, 16]. The systematic analysis based on clustered and homologous miRNAs is quite necessary to unveil the potential functional correlation and contribution in tumorigenesis.

In the present study, to further understand the potential expression and functional correlations between miRNAs, we performed a global analysis of miRNA gene clusters and families in breast cancer using small RNA deep sequencing datasets. These related miRNAs may have higher sequence similarity (homologous miRNAs) or may be expressed in a single polycistronic transcript with close physical distance on chromosome (clustered miRNAs). They have been identified as cooperative regulatory molecules via contributing

to multiple biological processes. Simultaneously, they also have close phylogenetic relationships through complex evolutionary process. Based on their functional and evolutionary relationships, the expression analysis will provide information of indirect interaction between miRNAs and potential contribution in cancer development.

2. Materials and Methods

2.1. Source Data. High-throughput miRNA sequencing datasets of 4 paired tumor (breast cancer) and adjacent normal tissues (P1, P5, P6, and P7) were obtained from Guo et al. [17]. The information on miRNA gene clusters and families was obtained from the public miRBase database (Release 19.0, <http://www.mirbase.org/>). Abundantly expressed miRNA gene clusters and families were collected and further analyzed according to relative expression levels. To comprehensively track the expression profiles between clustered or homologous miRNAs, we collected and analyzed all the members of miRNA clusters and families if one member was abundantly expressed in a sample.

2.2. Expression Analysis. The expression patterns were estimated using the relative expression levels (percentage) in every miRNA gene cluster or family. Simultaneously, due to dynamic expression across different individuals, equally mixed datasets were also used to estimate the expression patterns. We analyzed the potential relationships between miRNA gene clusters and families, especially some miRNAs could be yielded by multicopy pre-miRNAs. According to abundantly expressed miRNAs, we attempted to discover the potential cross-distribution and expression patterns between clustered miRNAs and homologous miRNAs. Moreover, we also focused on those clustered miRNAs and homologous miRNAs that were identified as sense and antisense miRNAs in the specific genome locus. Further expression analysis was performed based on the 4 paired datasets and mixed datasets, respectively.

2.3. Gene Ontology Enrichment Analysis. Experimentally validated target mRNAs of deregulated miRNAs were obtained from the miRTarBase database [18]. For those miRNAs with less or no validated targets, target mRNAs were predicted based on “seed sequences” using the TargetScan program [19]. According to these target mRNAs of deregulated miRNA gene clusters and families, the functional enrichment analysis was performed using CapitalBio Molecule Annotation System V4.0 (MAS, <http://bioinfo.capitalbio.com/mas3/>).

3. Results

Abundantly expressed clustered and homologous miRNAs were selected to perform further analysis. Some abundantly and abnormally expressed miRNAs (such as miR-23a, miR-23b, miR-24, miR-222, and miR-29a) had been experimentally validated using real-time PCR in breast cancer samples [20]. Interestingly, we found that many miRNA gene clusters

and families had close relationships or had overlapped members (Tables S1 and S2; see Supplementary Material available online at <http://dx.doi.org/10.1155/2014/782490>). Some miRNAs could be yielded by different pre-miRNAs, and the phenomenon of multicopy pre-miRNAs largely contributed to the complex relationships. Generally, these pre-miRNAs may be located on different chromosomes, different strands of the same chromosome (including sense and antisense strands), or different regions on the same strand. The various distributions complicated the compositions of miRNA gene clusters and families. For example, miR-221 and miR-222 were members of miR-221 gene family with higher sequence similarity, but they were also clustered on chromosome X and identified as miR-222 gene cluster. Homologous miRNA members could be located in different gene clusters through locating on different genomic regions or different chromosomes. For example, miR-23a and miR-27a were clustered on chromosome 19, while miR-23b and miR-27b were located in a cluster on chromosome 9. Simultaneously, sense and antisense miRNA genes were also involved in the gene cluster and family (Tables S1 and S2). miR-103a and miR-103b were homologous miRNA species (they were homologous members in miR-103 gene family), while their precursors were located on the sense and antisense strands of chromosomes 5 and 20, respectively (miR-103a-2 and miR-103a-1 gene clusters could be detected based on their multicopy pre-miRNAs).

Clustered and homologous miRNAs always showed consistent deregulation patterns in tumor samples (Figure 1(a)), although they had various expression levels (Figure 1(b)). They might show expression divergence as well as individual difference across different samples. The dynamic expression patterns in miRNA gene clusters and families were quite popular, even though they might be cotranscribed as a single polycistronic unit or had higher sequence similarity. For example, one member was abundantly expressed, while another clustered or homologous member had lower expression level (Figure 1(b)). The deregulation patterns were also influenced by the various expression levels, especially some were rarely expressed. The fold change (\log_2) showed larger divergence between different clustered or homologous miRNA species and between different individuals (Figure 1). Furthermore, we also performed the expression analysis based on the mixed datasets. Similar expression patterns could be detected (Figure 2). The divergence of fold change existed, but the difference had been largely reduced than the expression analysis based on each pair of samples (Figures 1 and 2).

For those miRNA gene clusters and families that were involved in sense and antisense miRNAs, we also analyzed their expression patterns. As expected, they always showed larger expression divergence (or both of them were rarely expressed): if one member had abundant expression level, another would be rarely detected (Figure 3). The sense and antisense miRNAs could be perfectly reverse complementarily binding to each other, although they may also be homologous miRNA genes with higher sequence similarity.

According to the predicted target mRNAs, the common targets could be detected between clustered or homologous

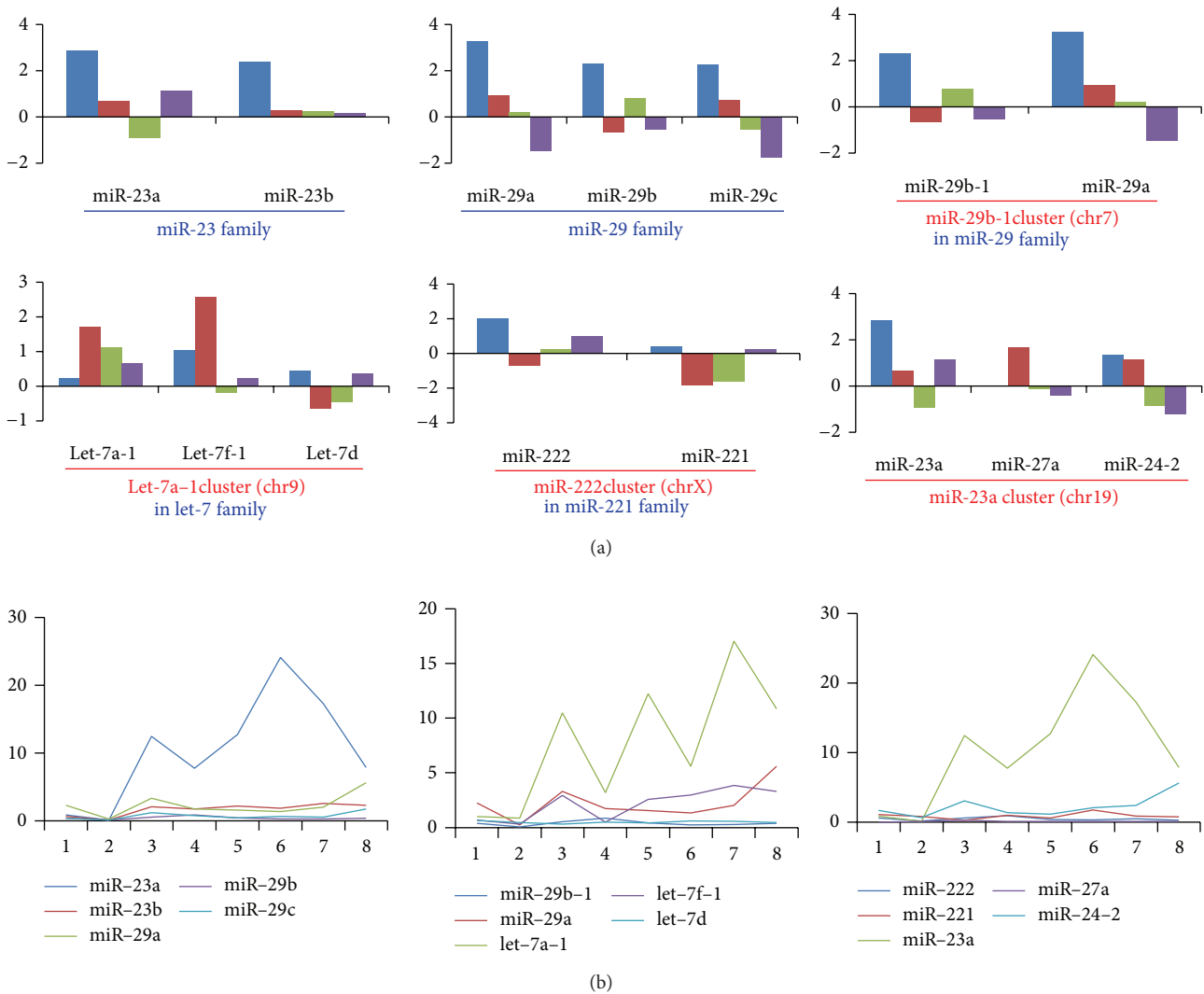


FIGURE 1: Examples of deregulation patterns of miRNA gene clusters and families (a) and their dynamic expression patterns (b). (a) Members in these miRNA gene clusters and families may be repeated. Some clusters are members of a specific gene family. The horizontal axis indicates the miRNA gene cluster or gene family; and the vertical axis indicates the fold change value (log₂) based on each pair of tumor and adjacent normal samples. Bars in different colors (blue, red, green, and purple) indicate fold change value (log₂) of the four pairs of tumor and adjacent normal samples, respectively; (b) dynamic expression across the 8 involved samples (P1-tumor, P1-normal, P5-tumor, P5-normal, P6-tumor, P6-normal, P7-tumor, and P7-normal). The horizontal axis indicates the 8 samples, and the vertical axis indicates the relative expression (percentage).

miRNAs (Table S3). Functional enrichment analysis of deregulated miRNA groups showed that they had versatile roles in multiple basic biological processes such as regulation of transcription and signal transduction (Table 1).

4. Discussion

miRNAs have been widely studied as crucial regulatory molecules, but the global expression patterns of miRNA gene clusters and families are little known. These clustered or homologous miRNAs have potential, functional, and evolutionary relationships, and they may coregulate or coordinately regulate multiple biological processes. The potential

coordinated interaction complicates the coding-non-coding RNA regulatory network and enriches the miRNA-mRNA and miRNA-miRNA interactions [21, 22]. Sense and antisense miRNAs have been characterized as potential miRNA-miRNA interaction with larger expression divergence (Figure 3). Recent studies have shown that these endogenous complementary miRNAs can restrict the transcription or maturation process of one another [23–27]. The perfectly reverse binding suggests that miRNA-miRNA interaction may be a potential regulatory method in the miRNA world [21]. Further, the compositions of gene clusters and families are not random and independent, and the phenomenon of multicopy pre-miRNAs further



FIGURE 2: Examples of consistent or similar deregulation patterns in clustered and homologous miRNAs based on the equally mixed datasets. The horizontal axis indicates the miRNAs and the involved gene cluster and family, and the vertical axis indicates the fold change (log₂) based on the equally mixed datasets of tumor and normal samples. The green and red bars indicate the threshold values (2 and -2).

TABLE 1: Enriched GO terms based on experimentally validated target mRNAs in Figure 1.

GO term	Count	P value
GO:0006355 regulation of transcription, DNA-dependent	24	5.28E - 26
GO:0006350 transcription	18	7.48E - 17
GO:0007165 signal transduction	18	2.37E - 14
GO:0007275 development	15	1.82E - 13
GO:0006508 proteolysis	14	3.32E - 17
GO:0045944 positive regulation of transcription from RNA polymerase II promoter	12	1.05E - 21
GO:0007155 cell adhesion	10	1.67E - 12
GO:0006915 apoptosis	10	5.62E - 12
GO:0008285 negative regulation of cell proliferation	9	7.05E - 15
GO:0006468 protein amino acid phosphorylation	9	2.03E - 11
GO:0006917 induction of apoptosis	8	9.59E - 14
GO:0042981 regulation of apoptosis	8	2.75E - 10

Here, we only list important GO terms that involved at least 8 target mRNAs of differentially expressed miRNAs. Count indicates involved number of target mRNAs; P value indicates enrichment P value.

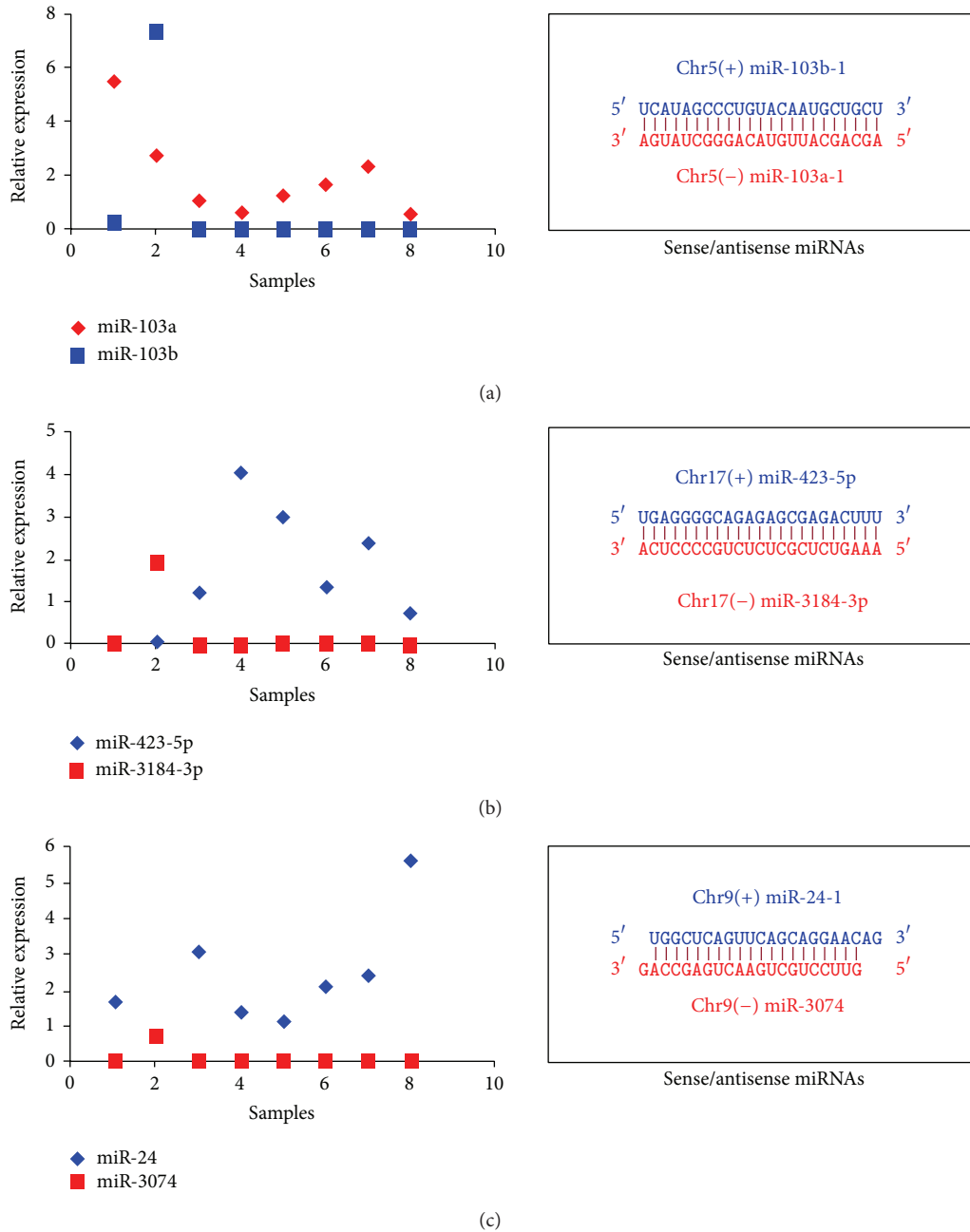


FIGURE 3: Examples of diverged expression of sense and antisense miRNAs in miRNA gene clusters and families. The detailed sense and antisense miRNA sequences are also presented on light, and they can perfectly reverse complementarily binding to each other. (a) miR-103a and miR-103b are homologous miRNAs and also clustered on chromosomes 5 and 20 as sense and antisense miRNAs (the figure only lists the sequences on chromosome 5); (b) miR-423 and miR-3184 are a pair of sense and antisense miRNAs on chromosome 17; (c) miR-24-1 and miR-3074 are sense and antisense miRNAs on chromosome 9 and also clustered in miR-23b gene cluster.

complicates the distributions of miRNAs [28]. Clustered and homologous miRNAs always have close relationships with overlapped members (Tables S1 and S2). The interesting distributions and relationships may be mainly derived from the complex duplication history that may adapt to the functional and evolutionary pressures [13–15, 29].

Although clustered and homologous miRNA members are involved in various and inconsistent enrichment levels

via maturation and degradation mechanisms, they are prone to present consistent or similar deregulation patterns in tumor samples (Figures 1 and 3). Across different samples, miRNAs may show the larger expression divergence. The reason may be partly derived from the deep sequencing datasets with higher sensitivity and potential divergence during sequencing and sample preparation. On the other hand, the individual difference also leads to the expression

divergence, especially for these patients may be involved in different degrees or stages of breast cancer, although they are clinically characterized as primary breast cancer. Multiple factors may contribute to occurrence and development of breast cancer, and different samples may be prone to detect slightly inconsistent miRNA expression profiles. The dynamic expression patterns may contribute to the robust regulatory network and adapt to specific intracellular environment. Indeed, these miRNA gene clusters and families have important roles in multiple biological processes (Table 1). The consistent deregulation patterns contribute to their potential coordinated interaction, although they indicate various expression levels.

Furthermore, other factors also contribute to the expression divergence in miRNA gene clusters and families. Firstly, the phenomenon of cross-mapping or multiple mapping contributes to the relative expression levels [23, 30], especially between those homologous miRNAs. The same sequencing fragments can be mapped to different pre-miRNA sequences, and any arbitrary selection will influence the final expression analysis. Secondly, multiple pre-miRNAs have been identified that can yield the same miRNAs. However, it is hard to infer the genuine origin. These multiple pre-miRNAs are always located on different chromosomes or different strands on the same chromosomes. In the typical analysis, we always analyze the mature miRNAs and rarely consider their real origins. The default analysis would influence the expression patterns of members in miRNA gene clusters. Clustered miRNAs are characterized based on the location distributions of miRNA genes, but mature miRNAs are used to estimate the final expression levels. The arbitrary and default selection may lead to the imprecise expression analysis. Finally, an miRNA locus can yield many sequences with various 5' and/or 3' ends due to imprecise cleavage of Drosha and Dicer [31–33]. These multiple miRNA variants, also termed isomiRs, largely enrich the miRNA study and coding-non-coding RNA regulatory network as physical miRNA isoforms. These multiple isomiRs also influence the expression estimation, especially expression analysis based on the most abundant isomiR, the canonical miRNA, or sum of all isomiRs, respectively. Simultaneously, these various sequences also contribute to the phenomenon of cross-mapping between different miRNAs [23]. In the present study, the expression analysis at the miRNA level (based on the sum of all isomiRs) is not comprehensive. Collectively, expression divergence between miRNAs is more complexity *in vivo*, which may contribute to the dynamic regulatory network.

Taken together, although various expression levels can be detected, consistent or similar deregulation patterns are always found between clustered or homologous miRNAs. The expression patterns provide an opportunity to coregulate or coordinately regulate biological processes. Therefore, the dynamic and coordinated expression may have important biological roles, which should be derived from the functional and evolutionary pressures. As flexible regulatory molecules, multiple miRNAs can negatively regulate biological pathways based on potential coordinated interaction (e.g., based on miRNA gene clusters and families). Further study should

be performed that clustered and/or homologous miRNAs would be potential biomarkers to study the mechanisms in tumorigenesis.

Conflict of Interests

The authors declare no potential conflict of interests with respect to the authorship and/or publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61301251, 81072389, 81373102, and 81102182), the Research Fund for the Doctoral Program of Higher Education of China (nos. 211323411002 and 20133234120009), the China Postdoctoral Science Foundation funded project (no. 2012M521100), the Key Grant of 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 (nos. 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 (PAPD) of Jiangsu Higher Education Institutions.

References

- [1] D. P. Bartel, "MicroRNAs: genomics, biogenesis, mechanism, and function," *Cell*, vol. 116, no. 2, pp. 281–297, 2004.
- [2] G. A. Calin, "A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia (vol 353, pg 1793, 2005)," *New England Journal of Medicine*, vol. 355, pp. 533–533, 2006.
- [3] C. Caldas and J. D. Brenton, "Sizing up miRNAs as cancer genes," *Nature Medicine*, vol. 11, no. 7, pp. 712–714, 2005.
- [4] G. A. Calin, C. Sevignani, C. D. Dumitru et al., "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2999–3004, 2004.
- [5] C. Swanton and C. Caldas, "Molecular classification of solid tumours: towards pathway-driven therapeutics," *British Journal of Cancer*, vol. 100, no. 10, pp. 1517–1522, 2009.
- [6] D. Madhavan, M. Zucknick, M. Wallwiener et al., "Circulating miRNAs as surrogate markers for circulating tumor cells and prognostic markers in metastatic breast cancer," *Clinical Cancer Research*, vol. 18, pp. 5972–5982, 2012.
- [7] M. Redova, J. Sana, and O. Slaby, "Circulating miRNAs as new blood-based biomarkers for solid cancers," *Future Oncology*, vol. 9, no. 3, pp. 387–402, 2013.
- [8] L. P. Lim, M. E. Glasner, S. Yekta, C. B. Burge, and D. P. Bartel, "Vertebrate microRNA genes," *Science*, vol. 299, no. 5612, p. 1540, 2003.
- [9] J. Z. Xu and C. W. Wong, "A computational screen for mouse signaling pathways targeted by microRNA clusters," *RNA*, vol. 14, no. 7, pp. 1276–1283, 2008.

- [10] J. Yu, F. Wang, G. Yang et al., "Human microRNA clusters: genomic organization and expression profile in leukemia cell lines," *Biochemical and Biophysical Research Communications*, vol. 349, no. 1, pp. 59–68, 2006.
- [11] L. Guo and Z. Lu, "Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data," *Computational Biology and Chemistry*, vol. 34, no. 3, pp. 165–171, 2010.
- [12] S. R. Viswanathan, C. H. Mermel, J. Lu, C. Lu, T. R. Golub, and G. Q. Daley, "MicroRNA expression during trophectoderm specification," *PLoS ONE*, vol. 4, no. 7, Article ID e6143, 2009.
- [13] J. Hertel, M. Lindemeyer, K. Missal et al., "The expansion of the metazoan microRNA repertoire," *BMC Genomics*, vol. 7, p. 25, 2006.
- [14] R. Zhang, Y. Peng, W. Wang, and B. Su, "Rapid evolution of an X-linked microRNA cluster in primates," *Genome Research*, vol. 17, no. 5, pp. 612–617, 2007.
- [15] L. Guo, B. Sun, F. Sang, W. Wang, and Z. Lu, "Haplotype distribution and evolutionary pattern of miR-17 and miR-124 families based on population analysis," *PLoS ONE*, vol. 4, no. 11, Article ID e7944, 2009.
- [16] L. Guo and Z. Lu, "The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule?" *PLoS ONE*, vol. 5, no. 6, Article ID e11387, 2010.
- [17] L. Guo, Y. Zhao, S. Yang, M. Cai, Q. Wu, and F. Chen, "Genome-wide screen for aberrantly expressed miRNAs reveals miRNA profile signature in breast cancer," *Molecular Biology Reports*, vol. 40, no. 3, pp. 2175–2186, 2013.
- [18] S.-D. Hsu, F. Lin, W. Wu et al., "MiRTarBase: a database curates experimentally validated microRNA-target interactions," *Nucleic Acids Research*, vol. 39, no. 1, pp. D163–D169, 2011.
- [19] B. P. Lewis, I.-H. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge, "Prediction of mammalian microRNA targets," *Cell*, vol. 115, no. 7, pp. 787–798, 2003.
- [20] Q. Wu, C. Wang, Z. Lu, L. Guo, and Q. Ge, "Analysis of serum genome-wide microRNAs for breast cancer detection," *Clinica Chimica Acta*, vol. 413, no. 13-14, pp. 1058–1065, 2012.
- [21] L. Guo, B. Sun, Q. Wu, S. Yang, and F. Chen, "miRNA-miRNA interaction implicates for potential mutual regulatory pattern," *Gene*, vol. 511, no. 2, pp. 187–194, 2012.
- [22] L. Guo, Y. Zhao, S. Yang, H. Zhang, and F. Chen, "Integrative analysis of miRNA-mRNA and miRNA-miRNA interactions," *BioMed Research International*, vol. 2014, Article ID 907420, 8 pages, 2014.
- [23] 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, no. 5, Article ID e20517, 2011.
- [24] K. E. Shearwin, B. P. Callen, and J. B. Egan, "Transcriptional interference—a crash course," *Trends in Genetics*, vol. 21, no. 6, pp. 339–345, 2005.
- [25] C. F. Hongay, P. L. Grisafi, T. Galitski, and G. R. Fink, "Antisense transcription controls cell fate in *Saccharomyces cerevisiae*," *Cell*, vol. 127, no. 4, pp. 735–745, 2006.
- [26] A. Stark, N. Bushati, C. H. Jan et al., "A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands," *Genes and Development*, vol. 22, no. 1, pp. 8–13, 2008.
- [27] E. C. Lai, C. Wiel, and G. M. Rubin, "Complementary miRNA pairs suggest a regulatory role for miRNA:miRNA duplexes," *RNA*, vol. 10, no. 2, pp. 171–175, 2004.
- [28] L. Guo, Y. Zhao, H. Zhang, S. Yang, and F. Chen, "Integrated evolutionary analysis of human miRNA gene clusters and families implicates evolutionary relationships," *Gene*, vol. 534, no. 1, pp. 24–32, 2014.
- [29] A. M. Heimberg, L. F. Sempere, V. N. Moy, P. C. J. Donoghue, and K. J. Peterson, "MicroRNAs and the advent of vertebrate morphological complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 8, pp. 2946–2950, 2008.
- [30] M. J. L. de Hoon, R. J. Taft, T. Hashimoto et al., "Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries," *Genome Research*, vol. 20, no. 2, pp. 257–264, 2010.
- [31] R. D. Morin, M. D. O'Connor, M. Griffith et al., "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells," *Genome Research*, vol. 18, pp. 610–621, 2008.
- [32] L. Guo, Q. Yang, J. Lu et al., "A comprehensive survey of miRNA repertoire and 3' addition events in the placentas of patients with pre-eclampsia from high-throughput sequencing," *PLoS ONE*, vol. 6, no. 6, Article ID e21072, 2011.
- [33] P. Landgraf, M. Rusu, R. Sheridan et al., "A Mammalian microRNA expression Atlas based on small RNA library sequencing," *Cell*, vol. 129, no. 7, pp. 1401–1414, 2007.