The oncogenic and prognostic potential of eight microRNAs identified by a synergetic regulatory network approach in lung cancer

Ramkrishna Mitra and

Department of Biomedical Informatics, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, USA, ramkrishna.mitra@vanderbilt.edu

Zhongming Zhao

Departments of Biomedical Informatics, Psychiatry, and Cancer Biology, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232, USA, zhongming.zhao@vanderbilt.edu

Abstract

Transcription factors (TFs) and microRNAs (miRNAs), the two main gene regulators in the biological system, control the gene expression at the transcriptional and post-transcriptional level, respectively. However, little is known regarding whether the miRNATF co-regulatory mechanisms, predicted by several studies, truly reflect the molecular interactions in cellular systems. To tackle this important issue, we developed an integrative framework by utilising four independent miRNA and matched mRNA expression profiling data sets to identify reproducible regulations, and demonstrated this approach in non-small cell lung cancer (NSCLC). Our analyses pinpointed several reproducible miRNA-TF co-regulatory networks in NSCLC from which we systematically prioritised eight hub miRNAs that may have strong oncogenic characteristics. Here, we discussed the major findings of our study and explored the oncogenic and prognostic potential of eight prioritised miRNAs through literature-mining based analysis and patient survival analysis. The findings provide additional insights into the miRNA-TF co-regulation in lung cancer.

Keywords

computational biology; lung cancer; reproducible regulation; oncogenic microRNAs; patient survival

1 Introduction

Transcription factors (TFs) are important regulators that either induce or repress gene expression by binding to a gene's promoter region at the transcriptional level. Then, the expression of the same gene can be repressed by microRNAs (miRNAs), the small (~21–23 nucleotides) endogenous non-coding RNA molecules (Bartel, 2004), at the post-transcriptional level. miRNAs can both regulate and be regulated by TFs, forming a

Correspondence to: Zhongming Zhao.

regulatory circuit (Gong et al., 2011; Guo et al., 2010). Hence, it is not surprising that, in gene regulatory networks both types of regulators are tightly related to each other.

The examination of regulatory networks showed that TFs and miRNAs can jointly regulate target gene expression and frequently form 3-node feed-forward loops (FFLs) (Figure 1) (Poos et al., 2013). Recently, Sun et al. extended the 3-node FFLs to 4-node and showed the oncogenic potential of the inferred 3-node and 4-node FFLs in glioblastoma multiforme (Sun et al., 2012). These regulatory loops serve as important motifs in gene regulatory networks and play critical roles in the development of cancer and other diseases, including breast cancer (Qin et al., 2014), colorectal cancer (Sengupta and Bandyopadhyay, 2013), glioblastoma multiforme (Setty et al., 2012; Sun et al., 2012), high-grade serous ovarian cancer (Zhao et al., 2013), osteosarcoma (Poos et al., 2013), schizophrenia (Guo et al., 2010), T-cell acute lymphoblastic leukaemia (Ye et al., 2012), and several others. It is imperative to mention here that these studies mostly rely on the predicted regulation information of miRNAs (Bandyopadhyay and Mitra, 2009; Betel et al., 2010; Friedman et al., 2009; Mitra and Bandyopadhyay, 2011) and TFs (Matys et al., 2006); hence, the identified regulatory networks may suffer from high false-positive molecular interactions.

Recently we uncovered the presence of miRNA-TF co-regulatory networks in non-small cell lung cancer (NSCLC) (Mitra et al., 2014). To minimise false-positives, we investigated the regulator-target relationships that were reproducible or preserved in multiple independent NSCLC data sets. Reproducible regulation has not yet been applied to miRNA-TF co-regulatory network analyses in cancer or other diseases, even though it may reflect the true molecular interaction (Dutta et al., 2012; Langfelder et al., 2011). The study constructed the miRNA-TF co-regulatory networks from a discovery data set and validated the regulations using three independent validation data sets where all the data sets consisted of matched miRNA and mRNA expression profiles in NSCLC.

Our network analysis prioritised eight miRNAs (miR-9-5p, miR-17-5p, miR-96-5p, miR-130b-3p, miR-182-5p, miR-183-5p, miR-200b-3p, and miR-200c-3p) and showed that they have strong oncogenic potential in NSCLC pathology. We reported that these miRNAs were up-regulated in NSCLC and a large proportion of their regulations were reproducible in other independent data sets. Importantly, we found that these eight miRNAs have the potential to repress the expression of 16 tumour suppressor genes and TFs in NSCLC and form a sub-network that included 32 edges. Among these 32 edges, 30 (93.7%) were reproducible in independent validation data sets.

Our framework appeared more robust in several respects when compared to other conventional network studies. The two leading advantages are:

- the framework may immediately pinpoint high-confident biological regulations based on the number of validation data sets validate the regulations

- we can prioritise densely connected network modules that have been enriched with reproducible edges.

For example, by using the discovery data set, we identified four miRNA-TF co-regulatory network modules that are potentially associated with the TGF-β signalling pathway.

Modules 1, 2, 3, and 4 consisted of 12.5%, 33.04%, 50.89%, and 88.89% reproducible network edges, respectively. This information correctly prompted us to conduct follow-up experiments on module 4 in order to validate the important regulations. Using a luciferase reporter assay, we confirmed that, in this module, the tumour suppressor gene *TGFBR2* is a direct target of miR-9-5p and miR-130b-3p. This confidence cannot be gained from existing network studies that rely on a single expression profiling data set.

To further explore the oncogenic and prognostic potential of these miRNAs prioritised by our computational approach, we performed additional analyses in this study. Firstly, we conducted an in-depth literature search to infer whether our prioritised miRNA pool dysregulated in other lung cancer clinical samples. This may clarify that the up-regulation of these miRNAs was not due to the artefacts of high-throughput expression profiling data sets that we used in our previous study. Secondly, literature-based evidence, along with an extended analysis using 170 high-quality NSCLC patient samples, revealed the importance of six miRNAs in lung cancer patient survival. In summary, the combined results of our previous (Mitra et al., 2014) and current study provided an adequate foundation for lung cancer investigators to conduct in-depth experiments to uncover the therapeutic potential of these miRNAs in the treatment of NSCLC.

2 Materials and methods

2.1 Literature search for dysregulated miRNAs in lung cancer

We collected studies that examine miRNA dysregulation in lung cancer by searching PubMed (http://www.ncbi.nlm.nih.gov/pubmed), using the keywords 'lung cancer' and 'microRNA'. We collected aberrantly expressed miRNAs in lung cancer from high-throughput (microarray and miRNA-Seq) and low-throughput [reverse transcriptase polymerase chain reaction (RT-PCR) and northern blot] experimental results published before 22 August, 2014. We initially read the titles and abstracts in order to exclude studies that mainly used bioinformatics approaches, reviews, comments, and those unrelated to lung cancer or miRNA. We further excluded the studies if miRNA dysregulation was observed due to the interference of drugs (including chemotherapy or radiotherapy) or genetic variations. We identified 130 papers that report dysregulated miRNAs in lung tumour tissues or lung cancer cell lines compared to normal lung tissue samples or cell lines.

2.2 Lung squamous cell carcinoma patient samples for survival analysis

To perform the survival analysis, we selected relevant patient samples and corresponding clinical information from Supplemental Tables S2 and S5 of Mitra et al., (2014). Additionally, for these patient samples, we downloaded the 'days to death' and 'days to last follow-up' information from The Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga/) in order to denote event and censor, respectively. Patients were censored from statistical analysis if they were alive and had five years of clinical follow-up. We obtained 170 patients, among them 12 who had neither the 'days to death' nor the 'days to last follow-up' information. Furthermore, 20 patients were excluded from the analysis because the reported event/censor information was not within the five year range. In total

138 patients were eligible for survival analysis; among those being eligible, mortality occurred in 55 patients (39.85%).

Cox regression (or proportional hazards regression) is a method for investigating the effect of several variables in the context of an outcome such as death for survival analysis. We performed univariate Cox hazard regression analysis based on miRNA expression profiles, stratified TNM stages [grouped by lower stages (stage I, IA, IB, IIA, and IIB) and higher stages (stage IIIA, IIIB, and IV)], stratified tumour sizes [grouped by smaller sizes (T1, T1a, T1b, T2, T2a, and T2b) and larger sizes (T3 and T4)], stratified lymphnode metastasis [grouped by no lymphnode metastasis (N0) and lymphnode metastasis (N1, N2, N3)], and the patient's age at the time of initial diagnosis. Patients were divided into a lower age group and a higher age group by calculating the median age (68 years). We performed the survival analysis using the R package Survival (Therneau and Grambsch, 2014).

3 Results and discussion

3.1 Reproducible regulatory networks prioritise most studied dysregulated miRNAs in lung cancer

In an on-going project, we conducted an in-depth literature search to identify global patterns of miRNA dysregulation in lung cancer (data not shown). We identified 413 miRNAs that had significant differential expression in lung tumour tissue samples or cell lines compared to normal lung tissue samples or cell lines. Among the eight miRNAs (miR-9-5p, miR-17-5p, miR-96-5p, miR-130b-3p, miR-182-5p, miR-183-5p, miR-200b-3p, and miR-200c-3p), except for miR-130b-3p, the elevated expression of the seven miRNAs had already been confirmed by RT-PCR in previous studies (Table 1). Combining the high-throughput and low-throughput experimental evidence (see Materials and methods), we found that most of the miRNAs had been previously reported by a small number of studies and hence researchers garnered less confidence about their pathogenic potential in lung cancer. Only 17 miRNAs out of 413, or 4.12%, had been reported by 10 or more studies; among them miR-182-5p (19 studies), miR-183-5p (13 studies), miR-200b-3p (12 studies), and miR-9-5p (10 studies) were prioritised through our network study (Mitra et al., 2014). Another two prioritised miRNAs, miR-96-5p (8 studies; ~top 5%) and miR-200c-3p (7 studies; ~top 7%) belong to the most studied miRNAs (Figure 2). These results provide an additional line of support that these miRNAs, prioritised by our network study, have significant aberrant expression in lung cancer and may play a major role in NSCLC development. Of note, the network analyses in our previous study were performed based on large-scale, genome-wide datasets (e.g., TCGA); thus, the results were not biased towards the specific report in literature. Furthermore, in our previous study using RT-PCR analysis, we confirmed the elevated expression of miR-130b-3p in NSCLC for the first time (Mitra et al., 2014).

3.2 Prognostic value of prioritised miRNAs in NSCLC

Among the eight potential oncogenic miRNAs, the elevated expression of miR-9-5p (Xu et al., 2014), miR-17-5p (Chen et al., 2013), miR-183 clustered miRNAs (miR-96-5p, miR-182-5p, and miR-183-5p) (Zhu et al., 2011), and miR-200c-3p (Tejero et al., 2014)

were associated with poor overall survival (OS) rates of NSCLC patients. With the exception of miR-17-5p, all of these studies conducted a survival analysis using NSCLC tissue samples. The correlation of miR-17-5p over-expression with the poor OS of patients with lung cancer was determined by serum samples. And, to the best of our knowledge, there is no report for miR-130b-3p and miR-200b-3p. Hence, here we investigated whether the elevated expression of miR-17-5p, miR-130b-3p, and miR- 200b-3p in NSCLC tissue samples, extracted from TCGA, was related to the prognosis of patients with lung cancer.

For each miR-17-5p, miR-130b-3p, and miR-200b-3p, we analysed the Kaplan-Meier survival curve as a first step in assessing the prognostic value of the corresponding miRNA in NSCLC. We then selected miR-17-5p and divided the patient samples into quartiles (Q1–Q4), according to the expression of the miRNA. Patients belonging to Q4 were selected for the high-expression group, and the rest of the patients (Q1–Q3) were selected for the low-expression group, as demonstrated in Nosho et al. (2014) and Yoo et al. (2009). In the Kaplan-Meier survival analysis, we observed a trend with patients with high miR-17-5p expression to have a poor OS, in comparison with those with low-miR-17-5p expression (see Figure 3). The median survival rate decreased to 16.2 months (35.3 vs. 19.1) when miR-17-5p registered at a high level. The univariate Cox hazard regression model analysis demonstrated that an elevated expression of miR-17-5p (P-value = 0.004, log-rank test, hazard ratio (HR) = 2.23; 95% confidence interval (CI) = 1.27 – 3.92) was associated with poor OS (see Table 2). Other variables, such as tumour stage (P-value = 0.07) and tumour size (P-value = 0.03) showed a significant association with poor OS; however, the effect was not as strong as that of miR-17-5p over-expression. A further point to be noted is that, in Kaplan-Meier survival analysis, an elevated expression of miR-130b-3p and miR-200b-3p was not observed to be associated with poor OS.

4 Conclusion

This extended analysis, combined with the results from Mitra et al., (2014), elucidates the fact that most of the prioritised miRNAs possess both oncogenic and prognostic potential in NSCLC. In summary, the concept of reproducible miRNA-TF co-regulatory networks may enhance our understanding of miRNA-TF co-regulatory mechanisms in cancer or other diseases. Due to the rapid growth of high-throughput expression profiling studies, we may access multiple miRNA and mRNA expression profiles for different diseases. Therefore, the computational framework we proposed is not only feasible to be applied to other disease studies, but also necessary for the identification of complex gene regulation in order to enhance our understanding of a specific disease of interest.

Acknowledgments

Biographies

Ramkrishna Mitra received his PhD in Engineering from Jadavpur University, Kolkata, India in 2011. Currently, he is a postdoctoral fellow in the Department of Biomedical Informatics at Vanderbilt University School of Medicine. His research interests include the study of cancer mechanisms at the systems biology level using high-throughput biological data, including microRNA and gene expression profiles, protein-protein interaction networks, microRNA and transcription factor co-regulatory networks, among others.

Zhongming Zhao received his PhD in Human and Molecular Genetics from the Graduate School of the University of Texas Health Science Center at Houston and MD Anderson Cancer Center, Houston, Texas in 2000. Currently, he is Ingram Professor of Cancer Research and Professor in the Departments of Biomedical Informatics, Psychiatry, and Cancer Biology at Vanderbilt University School of Medicine. His research interests include bioinformatics and systems biology approaches to studying complex diseases, personalised medicine, pharmacogenomics, and biomedical informatics.

References

Bandyopadhyay S, Mitra R. TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. Bioinformatics. 2009; 25:2625–2631. [PubMed: 19692556]

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116:281–297. [PubMed: 14744438]

Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. 2010; 11:R90. [PubMed: 20799968]

Chen Q, Si Q, Xiao S, Xie Q, Lin J, Wang C, Chen L, Wang L. Prognostic significance of serum miR-17-5p in lung cancer. Med. Oncol. 2013; 30:353. [PubMed: 23263848]

Dutta B, Pusztai L, Qi Y, Andre F, Lazar V, Bianchini G, Ueno N, Agarwal R, Wang B, Shiang CY, Hortobagyi GN, Mills GB, Symmans WF, Balazsi G. A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. Br. J. Cancer. 2012; 106:1107–1116. [PubMed: 22343619]

Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19:92–105. [PubMed: 18955434]

Gong X, Sun J, Zhao Z. Gene regulation in glioblastoma: a combinatorial analysis of microRNAs and transcription factors. Int. J. Comput. Biol. Drug Des. 2011; 4:111–126. [PubMed: 21712563]

Guo AY, Sun J, Jia P, Zhao Z. A novel microRNA and transcription factor mediated regulatory network in schizophrenia. BMC Syst. Biol. 2010; 4:10. [PubMed: 20156358]

Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014; 42:D68–D73. [PubMed: 24275495]

Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput. Biol. 2011; 7:e1001057. [PubMed: 21283776]

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–D110. [PubMed: 16381825]

Mitra R, Bandyopadhyay S. MultiMiTar: a novel multi objective optimization based miRNA-target prediction method. PLoS One. 2011; 6:e24583. [PubMed: 21949731]

Mitra R, Edmonds MD, Sun J, Zhao M, Yu H, Eischen CM, Zhao Z. Reproducible combinatorial regulatory networks elucidate novel oncogenic microRNAs in non-small cell lung cancer. RNA. 2014; 20:1356–1368. [PubMed: 25024357]

Nosho K, Igarashi H, Nojima M, Ito M, Maruyama R, Yoshii S, Naito T, Sukawa Y, Mikami M, Sumioka W, Yamamoto E, Kurokawa S, Adachi Y, Takahashi H, Okuda H, Kusumi T, Hosokawa M, Fujita M, Hasegawa T, Okita K, Hirata K, Suzuki H, Yamamoto H, Shinomura Y. Association of microRNA-31 with BRAF mutation, colorectal cancer survival and serrated pathway. Carcinogenesis. 2014; 35:776–783. [PubMed: 24242331]

Poos K, Smida J, Nathrath M, Maugg D, Baumhoer D, Korsching E. How microRNA and transcription factor co-regulatory networks affect osteosarcoma cell proliferation. PLoS Comput Biol. 2013; 9:e1003210. [PubMed: 24009496]

Qin S, Ma F, Chen L. Gene regulatory networks by transcription factors and microRNAs in breast cancer. Bioinformatics. 2014 [Epub ahead of print].

Russo F, Di Bella S, Nigita G, Macca V, Lagana A, Giugno R, Pulvirenti A, Ferro A. miRandola: extracellular circulating microRNAs database. PLoS One. 2012; 7:e47786. [PubMed: 23094086]

Sengupta D, Bandyopadhyay S. Topological patterns in microRNA-gene regulatory network: studies in colorectal and breast cancer. Mol. Biosyst. 2013; 9:1360–1371. [PubMed: 23475160]

Setty M, Helmy K, Khan AA, Silber J, Arvey A, Neezen F, Agius P, Huse JT, Holland EC, Leslie CS. Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. Mol. Syst. Biol. 2012; 8:605. [PubMed: 22929615]

Sun J, Gong X, Purow B, Zhao Z. Uncovering MicroRNA and transcription factor mediated regulatory networks in glioblastoma. PLoS Comput Biol. 2012; 8:e1002488. [PubMed: 22829753]

Tejero R, Navarro A, Campayo M, Vinolas N, Marrades RM, Cordeiro A, Ruiz-Martinez M, Santasusagna S, Molins L, Ramirez J, Monzo M. miR-141 and miR-200c as markers of overall survival in early stage non-small cell lung cancer adenocarcinoma. PLoS One. 2014; 9:e101899. [PubMed: 25003366]

Therneau, T.; Grambsch, P. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2014.

Xu T, Liu X, Han L, Shen H, Liu L, Shu Y. Up-regulation of miR-9 expression as a poor prognostic biomarker in patients with non-small cell lung cancer. Clin. Transl. Oncol. 2014; 16:469–475. [PubMed: 24019037]

Ye H, Liu X, Lv M, Wu Y, Kuang S, Gong J, Yuan P, Zhong Z, Li Q, Jia H, Sun J, Chen Z, Guo AY. MicroRNA and transcription factor co-regulatory network analysis reveals miR-19 inhibits CYLD in T-cell acute lymphoblastic leukemia. Nucleic Acids Res. 2012; 40:5201–5214. [PubMed: 22362744]

Yoo PS, Sullivan CA, Kiang S, Gao W, Uchio EM, Chung GG, Cha CH. Tissue microarray analysis of 560 patients with colorectal adenocarcinoma: high expression of HuR predicts poor survival. Ann. Surg. Oncol. 2009; 16:200–207. [PubMed: 19009247]

Zhao M, Sun J, Zhao Z. Synergetic regulatory networks mediated by oncogene-driven microRNAs and transcription factors in serous ovarian cancer. Mol. Biosyst. 2013; 9:3187–3198. [PubMed: 24129674]

Zhu W, Liu X, He J, Chen D, Hunag Y, Zhang YK. Overexpression of members of the microRNA-183 family is a risk factor for lung cancer: a case control study. BMC Cancer. 2011; 11:393. [PubMed: 21920043]
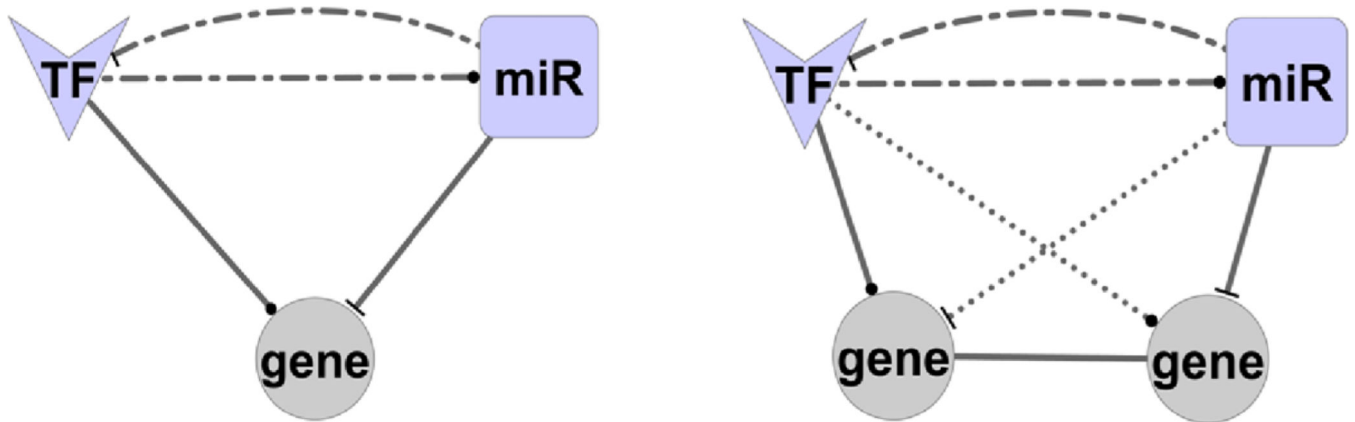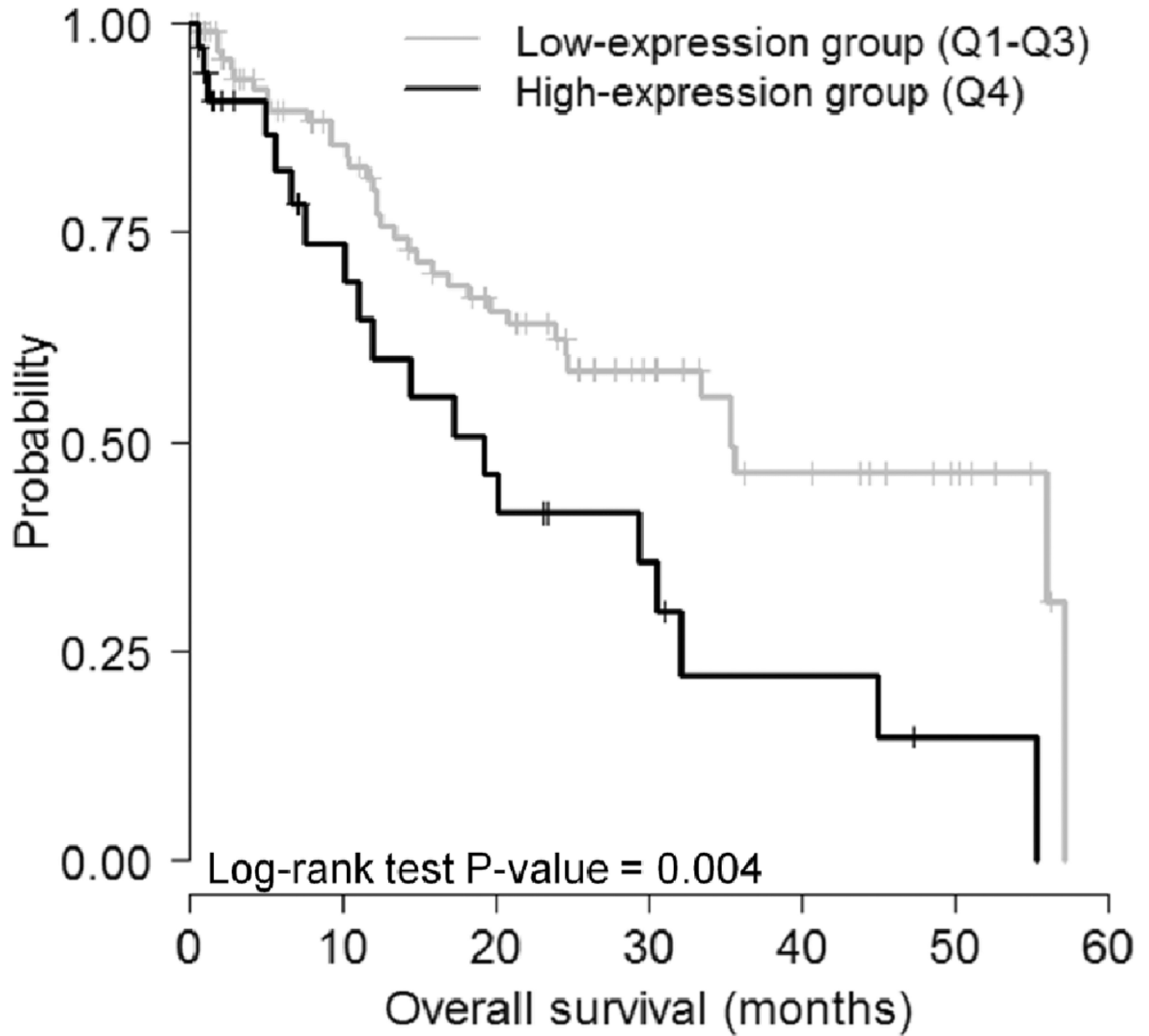
Figure 1.
microRNA- (miRNA) and transcription factor- (TF) mediated 3-node and 4-node feed-forward loops (FFLs). Solid line: required. Dash dot line: at least one of the dash dot lines is required. Round dot line: not required. 'T' shape arrowhead represents regulator-mediated target repression event. Round shape arrowhead represents regulator-mediated target activation or repression event. In 4-node FFLs, gene-gene association could be inferred from gene co-expression (Sun et al., 2012) or known protein-protein interaction (Poos et al., 2013) networks (see online version for colours)
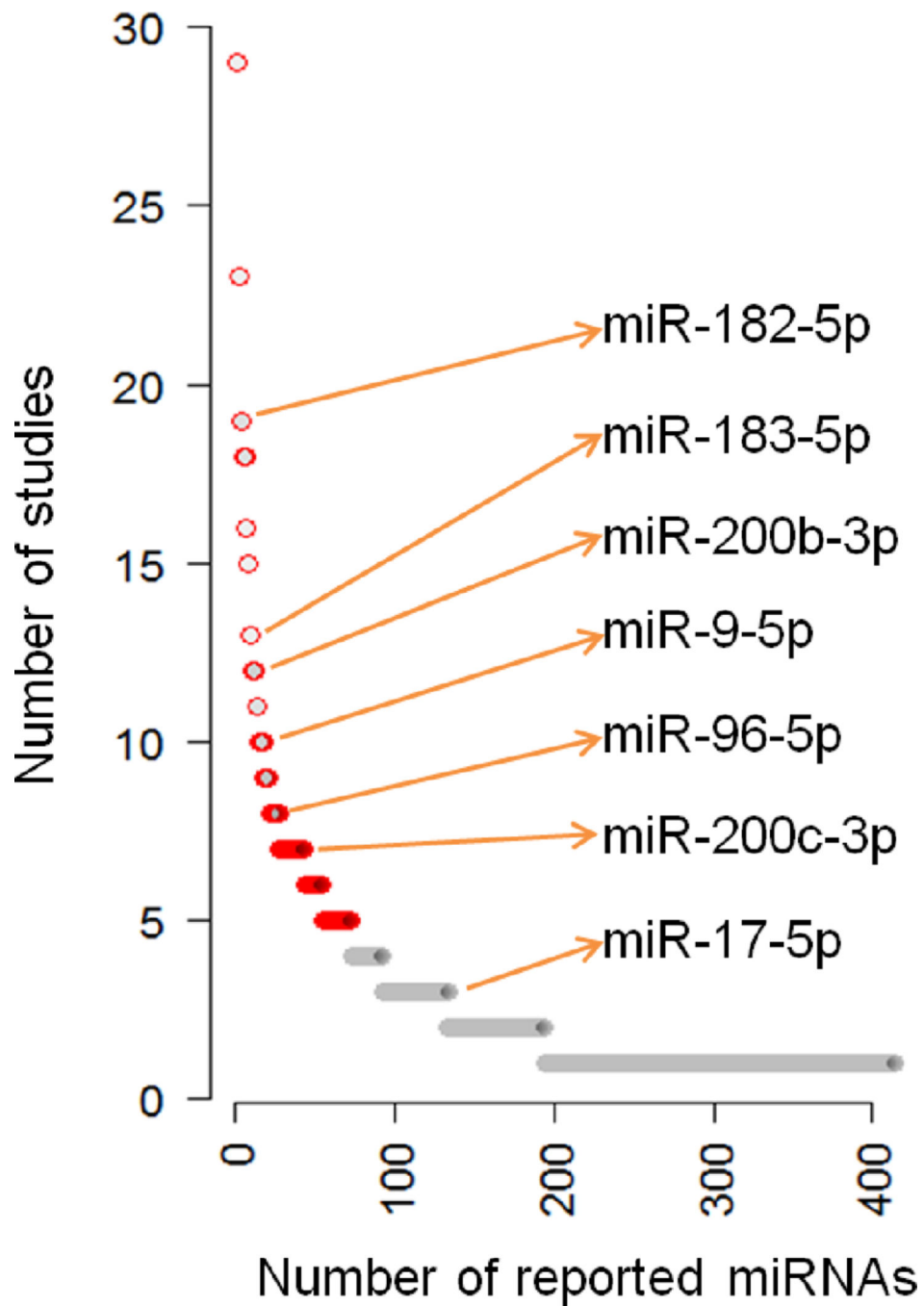
Figure 2.
Literature-based evidence of miRNAs found to be dysregulated in lung cancer by miRNA-TF co-regulation network approach. X-axis represents the number of miRNAs reported to be dysregulated in lung cancer in literature. Y-axis represents the number of studies for each miRNA. Red circle: miRNAs reported by 5 or more studies. Grey circle: miRNAs reported by less than five studies. miRNA IDs were converted according to miRBase v20 (Kozomara and Griffiths-Jones, 2014) using the ID conversion tool embedded in the web-server miRandola (Russo et al., 2012) (see online version for colours)

Figure 3.
Kaplan-Meier survival curves for patients with NSCLC plotted on miR-17-5p expression. Patients were categorised into low or high miR-17-5p expression group if they belong to the first three quartiles (Q1–Q3) or fourth quartile (Q4), respectively

Table 1

Dysregulation pattern of potential oncogenic miRNAs in lung cancer[§]

miRNA ID	No. studies reported up/down regulation	Total No. samples (tumour + normal) *tested*	Dysregulation confirmed by RT-PCR	PubMed IDs
miR-182-5p	19/0	1508	Up in eight studies	20885442, 19493678, 20526284, 21351266, 21358675, 21890451, 21748820, 21721011, 22573352, 22046296, 19584273, 19654003, 21116241, 21904633, 21516486, 21920043, 25012722, 24599520, 24519909
miR-183-5p	13/0	1028	Up in four studies	19493678, 20526284, 21358675, 21748820, 22573352, 19584273, 19654003, 21904633, 21516486, 21920043, 24599520, 24785186, 24113142
miR-200b-3p	12/0	680	Up in three studies	20885442, 19493678, 21351266, 21358675, 21563230, 19597153, 21300873, 21890451, 21721011, 21116241, 21516486, 24113142
miR-9-5p	8/2	993	Up in four studies	16778182, 21748820, 22573352, 19010987, 19654003, 21516486, 24113142, 24599520, 24019037, 24785186
miR-96	8/0	844	Up in three studies	20885442, 21563230, 22573352, 22046296, 21516486, 21920043, 24113142, 24599520
miR-200c-3p	7/0	455	Up in two studies	20885442, 19493678, 20526284, 21563230, 19584273, 21516486, 25124149
miR-17-5p	3/0	577	Up in two studies	19584273, 21544802, 19209007

[§]Using RT-PCR analysis, up-regulation of miR-130b-3p in NSCLC was confirmed by Mitra et al. (2014).

Table 2

Univariate Cox proportional hazard analysis of overall survival in NSCLC patients

Variable	HR	95% CI	P-value
miR-17-5p expression	2.23	1.27–3.92	0.004
TNM stage	1.69	0.94–3.02	0.07
Tumour size	2.04	1.05–3.98	0.03
Lymph node metastasis	1.48	0.86–2.56	0.16
Age	1.21	0.70–2.08	0.49

HR: Hazard ratio. CI: confidence interval.