Research paper

A novel framework for inferring condition-specific TF and miRNA co-regulation of protein–protein interactions

Junpeng Zhang a,* , Thuc Duy Le b , Lin Liu b , Jianfeng He c , Jiuyong Li b ,*  

a School of Engineering, Dali University, Dali, Yunnan 671003, China  
b School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia  
c School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan, 650500, China

Abstract

Recent studies have shown that transcription factors (TFs) and microRNAs (miRNAs), while independently regulate their downstream targets, collaborate with each other to regulate gene expression. However, their synergistic roles in protein–protein interactions (PPIs) remain mostly unknown. In this paper, we present a novel framework (called CoRePPI) for inferring TF and miRNA co-regulation of PPIs. Particularly, CoRePPI is aimed at discovering the co-regulation specific to a condition of interest, by using heterogeneous data, including miRNA and messenger RNA (mRNA) expression profiles, putative miRNA targets, TF targets and PPIs. CoRePPI firstly finds the network motifs indicating the co-regulation of PPIs by TFs and miRNAs in tumor and normal conditions separately. Then by identifying the differential motifs found in one condition but not in the other, it builds the networks consisting of TFs, miRNAs and their co-regulated PPIs specific to different conditions respectively. To validate CoRePPI, we apply it to the Pan-Cancer dataset which includes the expression profiles of 12 cancer types from TCGA. Through network topology analysis, we found that the tumor and normal CoRePPI networks are scale-free. Furthermore, the results of differential and intersected network analysis between the tumor and normal CoRePPI networks suggest that only a small fraction of the regulatory relationships between TFs and miRNAs are conserved in both conditions but they co-regulate different downstream PPIs in tumor and normal conditions; and in different conditions the majority of the regulatory relationships between TFs and miRNAs are different although they may regulate the same PPIs in their respective conditions. The CoRePPI sub-networks constructed for the three types of cancers (breast cancer, lung cancer and ovarian cancer) are all scale-free, and the intersection of these CoRePPI sub-networks can be utilized as the biomarker CoRePPI sub-network of the three types of cancers. The PPI enrichment analyses of the tumor and normal CoRePPI networks suggest that the co-regulating TFs and miRNAs are significantly associated with the specific biological processes, diseases and pathways. In addition, comparing with the two non-condition-specific approaches, the tumor CoRePPI network is found to have the most enriched cancer-related PPIs. Altogether, the results uncover the combined regulatory patterns of TFs and miRNAs on the PPIs, and may provide new insights for research in cancer-associated TFs and miRNAs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In order to understand the complex mechanisms of gene regulation, it is vital to study the regulation at both transcriptional and post-transcriptional levels.

At the transcriptional level, transcription factors (TFs) are the main regulators. They are proteins that bind to the enhancer or promoter regions of DNA adjacent to the target genes that they regulate. Depending on the TFs, the transcription of their target genes are either up or down regulated (Shen-Orr et al., 2002; Yu and Gerstein, 2006; Wang et al., 2011; Qi and Michoel, 2012; Rhee et al., 2014; Dailey, 2015). TFs have been found in all living organisms (van Nimwegen, 2003), and they are essential for many important biological processes, including development (Lobe, 1992), intercellular signaling (Pawson, 1993), and...
cell cycle (Wheaton et al., 1996). Many TFs are known to act as onco-
genes or tumor suppressors, thus mutations or aberrant regulation of
the TFs is associated with human cancers (Tupler and Green, 1999).
Unlike TFs, microRNAs (miRNAs) are small non-coding RNAs of ~22
nucleotides in length, and they act as the major post-transcriptional
regulators. miRNAs bind to the complementary sequences within their
target messenger RNAs (mRNAs), usually resulting in translational
repression or target degradation and gene silencing (Kusenda et al.,
2006; Bartel, 2009). Previous research (Krol et al., 2010; Shukla et al.,
2011; Takasaki, 2015) has revealed the roles of miRNAs in negative
regulation as well as their possible involvement in positive regulation.
By regulating their target genes, miRNAs are likely to control many
biological processes, including developmental timing, cell proliferation,
metabolism, differentiation, apoptosis, cellular signaling, stress
responses, and cancer development and progression (Ambros, 2001;
Ambros, 2003; Bartel, 2004; Cui et al., 2006; Du and Zamore, 2007;
Cheng et al., 2013; Lin and Gregory, 2015).

Given the important roles of TFs and miRNAs, it is of great interest to
have the unified gene regulatory networks involving both TFs and
miRNAs for understanding biological processes. Some studies (Shalgi
et al., 2007; Cui et al., 2007a; Le Béchec et al., 2011; Roqueiro et al.,
2012; Le et al., 2013; Guo et al., 2014; Hamed et al., 2015) have
investigated TF and miRNA co-regulatory networks, but they have not
considered the downstream protein–protein interactions (PPIs),
which are indeed critical to most biological processes. PPIs are at the
core of the entire interactomics system of all living cells, and PPI
networks can provide insight into the causes of diseases (Sanz-
Pamplona et al., 2012). Consequently, inferring the impact of upstream
TF and miRNA regulation on PPIs facilitates the understanding of
biological mechanisms within living cells.

There are several methods of identifying miRNA-regulated PPIs
(Liang and Li, 2007; Hsu et al., 2008; Yuan et al., 2009; Sass et al.,
2011; Tacutu et al., 2010; Yang et al., 2014), and they can be divided
into three categories (Zhu and Chen, 2014): (1) examining the correla-
tion between miRNAs and PPI networks (Liang and Li, 2007; Hsu et al.,
2008), (2) investigating miRNA co-regulated PPI networks (Yuan et al.,
2009; Sass et al., 2011), and (3) identifying miRNA-regulated PPI
networks in relation to specific diseases (Tacutu et al., 2010; Yang
et al., 2014). Methods in the first category look for positive correlation
between the number of miRNA target site types and protein connectiv-
ity (Liang and Li, 2007), and it was found that miRNA regulated proteins
had short distance and higher modularity than randomly selected
proteins (Hsu et al., 2008). Methods in the second category study how
miRNAs co-regulate PPIs. In (Yuan et al., 2009), it was discovered that
miRNA in the same clusters were more likely to be synergistic to regu-
late PPI networks, and in (Sass et al., 2011), it was found that miRNAs
coordinated to regulate protein complexes at the post-transcriptional
level. In the third category, different methods were proposed to identify
miRNA-regulated PPI networks in special diseases, such as aging related
diseases (Tacutu et al., 2010), gastric and breast cancers (Yang et al.,
2014). These works provide fundamental knowledge about the relationships
between miRNAs, PPIs and diseases. However, the co-regulation by TFs and
miRNAs on PPIs is still unclear.

So far, the investigation of the collaboration between TFs and
miRNAs in co-regulating PPIs has only just emerged. Lin et al. (Lin
et al., 2012) analyzed four types of motifs (single-regulation, co-
regulation, crosstalk and independent motifs) to find out the regulatory
patterns of TFs and miRNAs on protein interactome. Apart from the
single-regulation motif, other three motifs involve two regulators,
TF–TF, miRNA–TF and miRNA–miRNA respectively. The results suggest
that crosstalk motifs (consisting of the regulators and their non-
shared target genes) may have important downstream effects on sever-
al biological processes in living cells via regulating corresponding PPIs.

The above work, however, has several limitations. Firstly, the meth-
od in (Lin et al., 2012) uses target information that is based on sequence
data to discover the correlations between upstream regulators and their
downstream PPIs. Sequence data are static and thus the networks dis-
covered are also static in all biological conditions. In reality, gene regu-
latory mechanisms are dynamic and condition-specific (Zacher et al.,
2012). Therefore the static networks may not be very useful in revealing
the causes of diseases. Secondly, target prediction based on sequence
complementarity and/or structural stability of the putative duplexes
has a high rate of false discoveries (Rajewsky, 2006). Finally, the work
does not consider the regulatory relationships between TFs and
miRNAs, which, however, play important roles in feed-forward loop
(FFL) motifs and feedback loop (FLB) motifs (Zhang et al., 2015).

To address the above limitations, in this paper, we propose a novel
framework, CoRePPI (Condition-specific Regulation of PPIs) to infer
condition-specific TF and miRNA co-regulation of PPIs. Gene expression
data are used to deduce the dynamic and condition related mechanisms
of gene regulation. We also incorporate PPIs into a TF and miRNA co-
regulatory network and infer TF and miRNA co-regulation of PPIs in
a specific condition.

By making use of the complementary information of dynamic
(miRNA and mRNA expression profiles) and static (putative miRNA tar-
gets, TF targets and PPIs) data, the CoRePPI framework is able to identify
TF and miRNA co-regulation of PPIs that is specific to a condition of
interest (e.g. tumor) to reveal causes of diseases.

To validate the proposed framework, we apply it to the expression
profiles of 12 cancer types retrieved from The Cancer Genome Atlas
(TCGA). From the CoRePPI networks constructed using the expression
profiles (together with TF and miRNA targets, and PPIs), a number of
novel and interesting findings are obtained, which may provide insights
into the co-regulation of PPIs by TFs and miRNAs in the conditions of
interest.

2. Materials and methods

2.1. Data sources

The CoRePPI framework makes use of heterogeneous data, including
the expression profiles of TFs, miRNAs and mRNAs in tumor and normal
conditions, TF and miRNA target information, and PPIs.

The miRNA and mRNA expression profiles of TCGA (The Cancer Ge-
nome Atlas) Pan-Cancer (Cancer Genome Atlas Research Network et al.,
2013) are available under Synapse ID syn300013 with doi: 10.7303/
syn300013 (https://www.synapse.org/#!Synapse:syn300013). The
dataset contains 12 cancer types, including bladder urothelial carcinoma
(BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma
(COAD), glioblastoma multiforme (GBM), head and neck squamous cell
carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), acute mye-
loyd leukemia (LAML), lung adenocarcinoma (LUAD), lung squamous
cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), rectum
adenocarcinoma (READ), and uterine corpus endometrioid carcinoma
(UCEC). In total 2066 samples, including 1838 tumor samples and 228
normal samples, are used for this work.

We extract TF expression profiles from the mRNA expression
profiles of TCGA using the list of TF repertoire (Vaquerizas et al.,
2009). Putative target binding information of TF–miRNA is downloaded from
MIR@NT@N v1.2.1 (Le Béchec et al., 2011), and we use both MIR@NT@N v1.2.1 and CRSD (Liu et al., 2006) to obtain TF–TF and TF–miRNA
target information. In the CRSD database, TRANSFAC 9.3 (Matys et al.,
2003) and promoter databases (Halees and Weng, 2004) are integrated.
The miRNA target binding information is retrieved from TargetScan v6.2
(Lewis et al., 2005). The PPIs are obtained from HPRD v9 (Human
Protein Reference Database) (Prasad et al., 2009). We also use another
PPI dataset, the manually curated human signaling PPI network v6 in
(Cui et al., 2007b), for the assessment of the robustness of CoRePPI, i.e.
to test whether consistent results can be achieved as those with using
the HPRD PPIs.

To obtain cancer-related PPIs (for analyzing the results), we collect
two lists of cancer genes. The first list is created by combining the
To infer specific cancer CoRePPI sub-networks (i.e. the work presented in Section 3.4), we create a list of miRNAs and genes related to the three types of cancers (breast cancer, lung cancer and ovarian cancer) from a wide range of sources of experimentally validated results. The list of miRNAs are obtained by integrating the collections from HMDD v2.0 (Lu et al., 2008), miR2Disease (Jiang et al., 2009), miRCancer (Xie et al., 2013), oncomiRDB (Wang et al., 2014) and phenomiR v2.0 (Ruepp et al., 2010). The breast cancer genes are collected from COSMIC v74 (Futreal et al., 2004), GAD (Becker et al., 2004), OMIM (Hamosh et al., 2005), BCGD (Baasiri et al., 1999) and G2SBC (Mosca et al., 2010). The lung cancer genes are collected from COSMIC v74 (Futreal et al., 2004), GAD (Becker et al., 2004), OMIM (Hamosh et al., 2005), BCGD (Baasiri et al., 1999) and G2SBC (Mosca et al., 2010).

Fig. 1. A flowchart of the proposed CoRePPI framework for inferring and analyzing condition-specific CoRePPI networks. TF target information, miRNA target predictions, protein–protein interactions and expression profiles of TFs, miRNAs and mRNAs are used. In Step 1, differential expression analysis and sample splitting are conducted to generate differential expression profiles in tumor and normal conditions, respectively. In Step 2, for each condition we identify the TF and miRNA co-regulatory network with Pearson correlation analysis by combining putative target binding information. In Step 3, we identify condition-specific CoRePPI motifs that are specific to tumor and normal conditions, respectively. For each condition, we merge the condition-specific motifs to form the CoRePPI network for the condition. In Step 4, network topological analysis and PPI enrichment analysis are applied to the CoRePPI networks in tumor and normal conditions respectively.
In this section, we present the CoRePPI framework for identifying condition-specific TF and miRNA co-regulation of PPIs. As shown in Fig. 1, the overall process contains the following steps.

1. Data preparation. We firstly collect expression profiles, TF targets, miRNA targets and PPIs. After differential expression analysis and sample splitting, we obtain the differential expression profiles of TFs, miRNAs and mRNAs in tumor and normal conditions, respectively.

2. Inferring TF and miRNA co-regulatory networks. In this step, CoRePPI builds the co-regulatory networks in tumor and normal conditions respectively. By using the expression profiles of TFs, miRNAs and mRNAs in a specific condition, Pearson correlation coefficient of each pair of TF-TF, TF-miRNA, TF-mRNA, miRNA-TF and miRNA-mRNA is calculated. We then check the pairs against TF and miRNA target information, and remove the pairs that are not supported by the target information. In order to infer significant TF-TF, TF-miRNA, TF-mRNA, miRNA-TF and miRNA-mRNA interactions and to reduce false negatives (type II error), we rank the interactions (correlated pairs) based on the absolute values of their Pearson correlation coefficients, and only the interactions whose correlation coefficients (absolute values) are above the median (of all the absolute values) are chosen to be included in the TF and miRNA co-regulatory networks in tumor and normal conditions, respectively. In each condition, we then augment the TF and miRNA co-regulatory network with the PPIs.

3. Identifying CoRePPI motifs and building CoRePPI networks. In each condition, we firstly search the TF and miRNA co-regulatory network (augmented with the PPIs) for the 6 types of CoRePPI motifs (see next section for details). We extract differential or condition-specific CoRePPI motifs, i.e. the motifs that appear in the network of the specific condition only. Finally, merging the extracted differential motifs in each condition separately gives us the condition-specific CoRePPI networks.

4. Analyzing CoRePPI networks. In this step, network topology analysis and PPI enrichment analysis are conducted to study the condition-specific CoRePPI networks.

5. In the following, we will describe the key steps, Step (3) and Step (4) in detail.

2.3. Identifying CoRePPI motifs and building CoRePPI networks

As shown in Fig. 2, we have identified 6 types of CoRePPI motifs by considering the co-regulation of PPIs by TFs and miRNAs. Here, the co-regulation by a TF and a miRNA is determined by whether the two regulators share a common PPI. This means that the TF and miRNA may share a common target that is a party of a PPI (motifs 1 and 2 in Fig. 2); or the TF and miRNA each regulates a different party of the same PPI (motifs 3 to 6 in Fig. 2). In the figure, the interaction between a TF and a miRNA can be in any direction or in both directions (from TF to miRNA, from miRNA to TF or in both directions). All the 6 types of CoRePPI motifs contain FBL motifs between TFs and miRNAs, and motifs 1 and 2 include FFL motifs directly and motifs 3 to 6 contain FFL motifs indirectly.

NetMatch (Ferro et al., 2007) allows the searching of biological networks for sub-components matching a given query. Thus, the NetMatch plugin in Cytoscape (Shannon et al., 2003) is used to find CoRePPI motifs in a TF and miRNA co-regulatory network (augmented with PPIs) for a specific condition. After obtaining the motifs from the co-regulatory networks in tumor and normal conditions respectively, for each condition, we remove the common CoRePPI motifs between the two conditions to get the motifs specific to the condition. At last, in each condition we merge the condition-specific CoRePPI motifs to generate the CoRePPI network specific to that condition.

2.4. Analyzing CoRePPI networks

To understand the network topological properties of a condition-specific CoRePPI network, we use the NetworkAnalyzer plugin (Assenov et al., 2008) in Cytoscape to analyze the topology of the
network, specifically the node degree distribution of the network. The degree of a node is the number of edges connected with the node (including both incoming and outgoing edges). Node degree distribution $p(k)$ is defined to be the fraction of nodes in the network with a degree of $k$ ($k = 0, 1, 2, ...$). By fitting a power law curve to node degree distribution values, the dependency between a node degree and the number of nodes with the degree can be visualized. NetworkAnalyzer considers only data points with positive values for fitting the power law curve of the form $y = bx^a$. The $R^2$ value is a statistical measure of the linearity of the curve fit and is used to quantify the fit to the power line. The maximal value of $R^2$ is 1, and the larger the $R^2$ value is, the better the fit is. If a network whose node degree distribution follows a power law, at least asymptotically, the network is scale-free, which is one of most important characteristics of true complex biological networks (Barabasi and Oltvai, 2004). For the analysis, we treat the edges of a condition-specific CoRePPI network as undirected because in the analysis the degree of a node is defined as the number of edges connected with the node, without considering the directions of the edges.

To systematically analyze the CoRePPI networks, we categorize the regulatory relationships between TFs and miRNAs in the networks into two groups: (1) differential regulatory relationships between TFs and miRNAs, which are the TF-miRNA or miRNA-TF regulatory relationships only exist in one CoRePPI network, i.e. either in the tumor or normal CoRePPI network, but not in both; and (2) rewired regulatory relationships between TFs and miRNAs, which are the TF-miRNA or miRNA-TF regulatory relationships that are conserved or exist in both the tumor and normal CoRePPI networks, but such a TF-miRNA or miRNA-TF pair co-regulates different downstream PPI(s) in the two different conditions. To obtain the differential regulatory relationships between TFs and miRNAs, we extract the differential network from the tumor and normal CoRePPI networks, and the miRNA-TF and TF-miRNA regulatory relationships in the differential network are the above described differential regulatory relationships between TFs and miRNAs. To identify the rewired regulatory relationships between TFs and miRNAs, we extract the intersection of the tumor and normal CoRePPI networks, and the miRNA-TF and TF-miRNA regulatory relationships in the differential network are the above described rewired regulatory relationships between TFs and miRNAs.

To understand the underlying biological processes and pathways of downstream PPIs, we use the DAVID (Huang da et al., 2009) online tool (http://david.abcc.ncifcrf.gov/) to conduct PPI enrichment analysis, and we are only interested in GO (Gene Ontology) (Ashburner et al., 2000) biological processes and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) pathways at the significant level (adjusted $p$-value < 0.001, adjusted by Benjamini method).

To further study the statistical significance of the condition-specific CoRePPI networks, we use the method from Lin et al. (Lin et al., 2012) to calculate cancer-related PPI enrichment $z$-score (standard score) of the networks (see details of the method in Supplemental Material 1). The larger the absolute value of the $z$-score is, the more significant the correlation between the two synergistic regulators (TF and miRNA) and cancer-related PPIs is.

3. Results

3.1. Data processing

With the TCGA Pan-Cancer dataset, a miRNA or mRNA which has missing values in more than 50% of the samples is excluded from this study. Other miRNAs or mRNAs with less than 50% missing values in the dataset are imputed using the $k$-nearest neighbors (KNN) algorithm implemented in the R-package impute. The log2 transformed counts are then normalized using the quantile normalization method implemented in the R-package limma (Ritchie et al., 2015) of Bioconductor. We summarize the miRNA expression values by taking the average expression values of the miRNAs with the same gene family names.

After the data normalization, differential expression analysis is performed to identify differentially expressed miRNAs and mRNAs between tumor and normal samples, again using the limma package. To have a manageable number of miRNAs and mRNAs for CoRePPI, we select different cutoffs of adjusted $p$-values (adjusted by Benjamini–Hochberg method) in the differential expression analysis. As a result of the analysis, we identify 119 miRNAs (adjusted $p$-value < 0.01) and 3281 mRNAs (adjusted $p$-value < 0.0001) to be differentially expressed. Furthermore, we extract 332 TFs from the differentially expressed miRNAs using the list of TF repertoire (Vaquerizas et al., 2009). Therefore, the input of the Pearson correlation analysis includes the expression data of 332 TFs, 119 miRNAs and 2949 mRNAs in 1838 tumor and 228 normal samples. The processed dataset can be found in Supplemental Material 2.

3.2. The tumor and normal CoRePPI networks are scale-free

By following the CoRePPI framework shown in Fig. 1 and using the processed gene expression data (together with the corresponding TF and miRNA target information and the PPIs), we construct the tumor and normal CoRePPI networks (see Supplemental Material 3). Then NetworkAnalyzer is used to analyze the topological properties of the networks. The distributions of node degrees of the tumor and normal CoRePPI networks both approximately follow power law distributions and their $R^2$ values are both greater than 0.85 (see Fig. 3). The results indicate that the tumor and normal CoRePPI networks are both scale-free, suggesting that the tumor and normal CoRePPI networks possess the important characteristics of true complex biological networks.

3.3. CoRePPI network analysis uncovers differential and rewired regulatory relationships between TFs and miRNAs

Since the relationships between TFs and miRNAs are important gene regulatory relationships in complex gene regulatory networks, as described in Section 2.4, we perform differential and intersected network analysis between the tumor and normal CoRePPI networks to identify the differential and rewired regulatory relationships between TFs and miRNAs. As a result, we identify 939 differential (91.34%) and 89 rewired (8.66%) regulatory relationships between TFs and miRNAs. Therefore only a small fraction (8.66%) of the regulatory relationships between TFs and miRNAs are conserved across the CoRePPI networks of both conditions, but the conserved TF-miRNA or miRNA-TF pairs co-regulate different downstream PPIs in different conditions. At the same time, most regulatory relationships (91.34%) between TFs and miRNAs choose to co-regulate downstream PPIs only in one specific condition. The details of the differential and rewired regulatory relationships between TFs and miRNAs can be seen in Supplemental Material 4.

3.4. Specific cancer CoRePPI sub-networks

In this section, we select three popular cancer types, including breast cancer, lung cancer and ovarian cancer, to investigate the CoRePPI sub-networks that are specific to these cancer types. We define that specific cancer-related interactions are those in which the two interactive parties are specific cancer-related miRNAs, genes or proteins (including TFs). Using breast cancer, lung cancer and ovarian cancer miRNAs, genes and proteins, we extract three CoRePPI sub-networks from the tumor and normal CoRePPI networks, one for each of the three cancer types. As shown in Table 1, the distributions of node degrees in these three cancer CoRePPI sub-networks approximately follow power law distributions, with $R^2 = 0.823, 0.857$ and 0.956, respectively. Thus, these three cancer CoRePPI sub-networks are scale-free, suggesting that most specific cancer-related PPIs are co-regulated by a few specific cancer-related TFs and miRNAs. The details of the three cancer CoRePPI sub-networks can be seen in Supplemental Material 5.
Next, we identify the common or conserved CoRePPI motifs in all the three cancer CoRePPI sub-networks, and combine these motifs found to generate the conserved CoRePPI sub-network across the three cancer types (see Fig. 4). The conserved CoRePPI sub-network may have more biological significances than individual biomarkers, because it contains network motifs, which show not only the individual cancer miRNAs and genes, but also their regulatory relationships. Thus, it can be used as biomarker CoRePPI sub-network of these three cancers. We also extract the conserved CoRePPI sub-networks consisting of the CoRePPI motifs existing in two of the three cancers. We also extract the conserved CoRePPI sub-networks consisting of the CoRePPI motifs existing in two of the three cancers, and these conserved CoRePPI sub-networks can be seen in Supplemental Material 5.

3.5. PPI enrichment analysis of the tumor and normal CoRePPI networks

In order to understand the underlying biological processes and pathways of downstream PPIs, we identify significant biological processes...
may be closely associated with the chemical reactions and pathways of the macromolecule metabolic process through co-regulating these downstream genes.

For the tumor and normal CoRePPI networks, the top KEGG pathway of them is the same which is the pathway named hsa05200: Pathway in cancer (see Table 2). The result shows that 139 and 149 genes of the tumor and normal CoRePPI networks are significantly involved in the pathways in cancer, respectively. It suggests that the synergistic regulators (TFs and miRNAs) are also significantly associated with cancer in the tumor and normal CoRePPI networks by co-regulating these downstream genes.

In summary, the PPI enrichment analysis results demonstrate that our proposed method is able to find biologically valid results. The detailed results of the GO and KEGG enrichment analysis for the tumor and normal CoRePPI networks can be seen in Supplemental Material 6.

3.6. Comparison with the non-condition-specific approaches

We evaluate the performance of the CoRePPI framework by comparing its results (i.e. the obtained tumor and normal CoRePPI networks) with the results of two non-condition-specific methods: the method that only uses putative TF targets, miRNA targets and PPIs to identify TF and miRNA co-regulation of PPIs (called Non_CoRePPI_A); and the method that combines expression profiles of miRNAs and mRNAs, putative TF targets, miRNA targets and PPIs to infer TF and miRNA co-regulation of PPIs (called Non_CoRePPI_B). With the two methods, we also use Pearson correlation analysis to identify correlated pairs of TF-TF, TF-miRNA, TF-mRNA, miRNA-TF and miRNA-mRNA, but the samples of the expression profiles are not split according to their conditions.

Specifically, Non_CoRePPI_A only uses putative interactions and is similar to the method of Lin et al. (Lin et al., 2012). However, we use different network motifs (i.e. those 6 CoRePPI motifs shown in Fig. 2) to infer TF and miRNA co-regulation of PPIs.

Next, we use the two lists of cancer genes to calculate cancer-related PPI enrichment z-scores for each method. The first list of cancer genes is created by combining the two cancer gene databases, COSMIC and GAD, and the second list of cancer genes is obtained from IPA.

We define that cancer-related PPIs are those in which the two interactive parties of a PPI are cancer-related proteins. We use the HPRD database as the population of PPIs to compute the cancer-related PPI enrichment z-scores. The number of PPIs in HPRD is 36,852, and the

### Table 2

The enrichment results of downstream PPIs in the tumor and normal CoRePPI network. Top 5 enriched GO biological processes and KEGG Pathways are extracted. DAVID online tool is used to make GO and KEGG enrichment analysis and the p-values are adjusted by Benjamini method.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>#Downstream proteins</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5 enriched GO biological processes in the tumor CoRePPI network</td>
<td>GO:0010604 – positive regulation of macromolecule metabolic process</td>
<td>307</td>
<td>4.15E – 85</td>
</tr>
<tr>
<td></td>
<td>GO:0031328 – positive regulation of cellular biosynthetic process</td>
<td>257</td>
<td>6.93E – 75</td>
</tr>
<tr>
<td></td>
<td>GO:0051173 – positive regulation of nitrogen compound metabolic process</td>
<td>247</td>
<td>3.79E – 74</td>
</tr>
<tr>
<td></td>
<td>GO:0010628 – positive regulation of gene expression</td>
<td>233</td>
<td>3.72E – 74</td>
</tr>
<tr>
<td></td>
<td>GO:0010537 – positive regulation of macromolecule biosynthetic process</td>
<td>249</td>
<td>2.99E – 74</td>
</tr>
<tr>
<td>Top 5 enriched KEGG Pathways in the tumor CoRePPI network</td>
<td>hsa05200:Pathways in cancer</td>
<td>139</td>
<td>1.01E – 35</td>
</tr>
<tr>
<td></td>
<td>hsa04722:Neurotrophin signaling pathway</td>
<td>72</td>
<td>6.28E – 28</td>
</tr>
<tr>
<td></td>
<td>hsa04110:Cell cycle</td>
<td>61</td>
<td>2.25E – 18</td>
</tr>
<tr>
<td></td>
<td>hsa05220:Chronic myeloid leukemia</td>
<td>44</td>
<td>7.12E – 17</td>
</tr>
<tr>
<td></td>
<td>hsa05215:Prostate cancer</td>
<td>48</td>
<td>1.31E – 16</td>
</tr>
<tr>
<td>Top 5 enriched GO biological processes in the normal CoRePPI network</td>
<td>GO:0010604 – positive regulation of macromolecule metabolic process</td>
<td>240</td>
<td>1.17E – 40</td>
</tr>
<tr>
<td></td>
<td>GO:0010941 – regulation of cell death</td>
<td>232</td>
<td>1.06E – 40</td>
</tr>
<tr>
<td></td>
<td>GO:0043067 – regulation of programmed cell death</td>
<td>230</td>
<td>4.36E – 40</td>
</tr>
<tr>
<td></td>
<td>GO:0043068 – regulation of apoptosis</td>
<td>228</td>
<td>6.58E – 40</td>
</tr>
<tr>
<td></td>
<td>GO:0006357 – regulation of transcription from RNA polymerase II promoter</td>
<td>212</td>
<td>6.63E – 39</td>
</tr>
<tr>
<td>Top 5 enriched KEGG Pathways in the normal CoRePPI network</td>
<td>hsa05200:Pathways in cancer</td>
<td>149</td>
<td>3.16E – 39</td>
</tr>
<tr>
<td></td>
<td>hsa05220:Chronic myeloid leukemia</td>
<td>51</td>
<td>2.84E – 22</td>
</tr>
<tr>
<td></td>
<td>hsa04722:Neurotrophin signaling pathway</td>
<td>67</td>
<td>8.42E – 22</td>
</tr>
<tr>
<td></td>
<td>hsa04520:Adherens junction</td>
<td>45</td>
<td>5.17E – 16</td>
</tr>
<tr>
<td></td>
<td>hsa04510:Focal adhesion</td>
<td>79</td>
<td>1.60E – 15</td>
</tr>
</tbody>
</table>
numbers of cancer-related PPIs using the first and second lists of cancer genes are 7637 and 28,546, respectively.

Since the tumor and normal CoRePPI networks are generated by combining differential CoRePPI motifs, as shown in Table 3, the number of PPIs is smaller than those of the non-condition-specific methods. However, the CoRePPI network in tumor condition has the highest cancer-related PPI enrichment z-score. The result demonstrates that the differential CoRePPI motifs in tumor condition may play significant roles in PPI networks, and they are more likely to be associated with cancer. Furthermore, the CoRePPI network in normal condition also has high cancer-related PPI enrichment z-score. This result can be explained that if differential CoRePPI motifs in normal condition are removed, some diseases may also be induced.

It is noted that Non_CoRePPI_A (which does not use gene expression data) has the lowest cancer-related PPI enrichment z-scores. This result suggests that using expression profiles can improve the discovery of cancer-related PPIs. The details of TF and miRNA co-regulation of PPIs for Non_CoRePPI_A and Non_CoRePPI_B can be seen in Supplemental Material 7.

3.7. Robustness of the CoRePPI framework

To assess the robustness of our framework, we use another manually curated human signaling PPI network v6 (Cui et al., 2007b) to validate the results obtained using the HPRD database. With the other PPI network in (Cui et al., 2007b), the distributions of node degrees of the tumor and normal CoRePPI networks also approximately follow power law distributions, and their $R^2$ value are greater than 0.85 too, indicating that the tumor and normal CoRePPI networks are both scale-free (as shown in Fig. S1, Supplemental Material 1).

We also perform differential and intersected network analysis between the tumor and normal CoRePPI networks to identify differential and rewired regulatory relationships between TFs and miRNAs. As a result, we identify 866 differential (93.02%) and 65 rewired (6.98%) regulatory relationships between TFs and miRNAs. The result also implies that most regulatory relationships between TFs and miRNAs choose to co-regulate downstream PPIs in different conditions (details in Supplemental Material 4).

The constructed CoRePPI sub-networks specific to the three cancer types (breast cancer, lung cancer and ovarian cancer) approximately follow power law distributions, with $R^2 = 0.851, 0.840$ and 0.716, respectively, and are regarded as three scale-free biological networks (see Table S1, Supplemental Material 1). The details of the three cancer CoRePPI sub-networks can be seen in Supplemental Material 5. The conserved CoRePPI sub-networks across the three cancer types are also identified (see Fig. S2, Supplemental Material 1). The conserved CoRePPI sub-networks across two cancers of the three cancer types can be seen in Supplemental Material 5.

The GO enrichment analysis results show that the top GO biological process of the downstream PPIs of the tumor and normal CoRePPI networks are significantly associated with the regulation of cell death and the positive regulation of molecular function, respectively (see Table S2, Supplemental Material 1). For the tumor and normal CoRePPI networks, the top KEGG pathway of them is also the same which is the pathway named hsa05200: Pathway in cancer (see Table S2, Supplemental Material 1). For the tumor and normal CoRePPI networks are significantly associated with the regulation of cell death and the positive regulation of molecular function, respectively (see Table S2, Supplemental Material 1). For the tumor and normal CoRePPI networks, the top KEGG pathway of them is also the same which is the pathway named hsa05200: Pathway in cancer (see Table S2, Supplemental Material 1). For the tumor and normal CoRePPI networks, the top KEGG pathway of them is also the same which is the pathway named hsa05200: Pathway in cancer (see Table S2, Supplemental Material 1).

The constructed CoRePPI sub-networks specific to the three cancer types (breast cancer, lung cancer and ovarian cancer) approximately follow power law distributions, with $R^2 = 0.851, 0.840$ and 0.716, respectively, and are regarded as three scale-free biological networks (see Table S1, Supplemental Material 1). The details of the three cancer CoRePPI sub-networks can be seen in Supplemental Material 5. The conserved CoRePPI sub-networks across the three cancer types are also identified (see Fig. S2, Supplemental Material 1). The conserved CoRePPI sub-networks across two cancers of the three cancer types can be seen in Supplemental Material 5.

The GO enrichment analysis results show that the top GO biological process of the downstream PPIs of the tumor and normal CoRePPI networks are significantly associated with the regulation of cell death and the positive regulation of molecular function, respectively (see Table S2, Supplemental Material 1). For the tumor and normal CoRePPI networks, the top KEGG pathway of them is also the same which is the pathway named hsa05200: Pathway in cancer (see Table S2, Supplemental Material 1). The PPI enrichment analysis results also demonstrate that our proposed method can discover results that are biologically valid. The detailed results of the GO and KEGG enrichment analysis for the tumor and normal CoRePPI networks can be seen in Supplemental Material 6.

The comparison results with the non-condition-specific approaches also show that the CoRePPI network in tumor condition has the highest cancer-related PPI enrichment z-score, suggesting that using expression profiles can improve the discovery of cancer-related PPIs (as shown in Table S3, Supplemental Material 1).

In summary, the above results are consistent with those obtained using the HPRD database, indicating that CoRePPI is robust in inferring condition-specific TF and miRNA co-regulation of PPIs.

4. Conclusion

TFs and miRNAs are two primary metazoan gene regulators. TFs are vital for the regulation of gene expression at the transcriptional level, and miRNAs play important roles at the post-transcriptional level by regulating their target genes. However, a unified picture of the co-regulation of PPIs by the two types of regulators still remains unearthed. Since TF and miRNA regulation can influence protein production of their target genes, investigating the correlations between PPIs and their upstream regulators (TFs and miRNAs) facilitates the understanding of biological mechanisms within living cells.

In this study, we have proposed the CoRePPI framework and have applied it to infer condition-specific TF and miRNA co-regulation of PPIs from heterogeneous data, including miRNA and mRNA expression profiles of 12 cancer types from TCGA, putative miRNA and TF targets, and PPIs.

The framework has three features for inferring condition-specific CoRePPI networks. Firstly, we combine expression profiles with putative miRNA targets, TF targets, and PPIs to explore TF and miRNA co-regulation of PPIs. This will fully utilize the complementary information from both static data (putative miRNA targets, TF targets, and PPIs) and dynamic or condition-specific data (expression profiles). Secondly, we consider not only TF-target and miRNA-target relationships, but also the interplays between TFs and miRNAs, as relationships between TFs and miRNAs play vital roles in FFL motifs and FBL motifs. Lastly and most importantly, we consider the difference of TF and miRNA co-regulation of PPIs between tumor and normal conditions, and use the differential CoRePPI motifs to generate a condition-specific CoRePPI network. Due to this, CoRePPI takes full advantage of condition information of expression profiles, and the tumor CoRePPI networks could reveal the effects of TFs and miRNAs on their downstream PPIs and show causes of cancer better than the non-condition-specific methods.

The results of the application of CoRePPI to the TCGA Pan-Cancer data suggest that the proposed framework can be a promising approach to identifying cancer-associated TF and miRNA co-regulation of PPIs, and hence can provide new insights for understanding the mechanisms and roles of TF and miRNA regulation in biological processes, including cancer development.

Conflict of interest statement

The authors declare no competing financial interests.

<p>| Table 3 | Comparison between CoRePPI (CoRePPI_Tumor and CoRePPI_Normal) and two non-condition-specific methods in terms of cancer-related PPI enrichment z-scores. |</p>
<table>
<thead>
<tr>
<th>Methods</th>
<th>#PPIs</th>
<th>#Cancer-related PPIs</th>
<th>Cancer-related PPI enrichment z-score</th>
<th>Cancer-related PPI enrichment z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non_CoRePPI_A</td>
<td>10,826</td>
<td>2621</td>
<td>8837</td>
<td>24.0724</td>
</tr>
<tr>
<td>Non_CoRePPI_B</td>
<td>7076</td>
<td>1916</td>
<td>6003</td>
<td>56.8247</td>
</tr>
<tr>
<td>CoRePPI_Tumor</td>
<td>4292</td>
<td>1331</td>
<td>3625</td>
<td>71.7371</td>
</tr>
<tr>
<td>CoRePPI_Normal</td>
<td>3735</td>
<td>1024</td>
<td>3170</td>
<td>50.6780</td>
</tr>
</tbody>
</table>

\(^a\) The list of cancer genes is taken by combining two cancer genes databases: COSMIC and GAD.

\(^b\) The list of cancer genes is obtained from IPA.
Acknowledgments

This work has been partially supported by the Applied Basic Research Foundation of Science and Technology of Yunnan Province (No: 2013FD038), and the Australian Research Council Discovery Grant DP130104090.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jgene.2015.11.023.

References


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


catalogue. Nucleic Acids Res. 27, 297–300.


