



LAcModule: Identification of Competing Endogenous RNA Modules by Integrating Dynamic Correlation

Xiao Wen, Lin Gao* and Yuxuan Hu

School of Computer Science and Technology, Xidian University, Xi'an, China

OPEN ACCESS

Edited by:

Jie Sun,
Wenzhou Medical University, China

Reviewed by:

Quan Zou,
University of Electronic Science and
Technology of China, China
Fengbiao Mao,
University of Michigan, United States

*Correspondence:

Lin Gao
lgao@mail.xidian.edu.cn

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 11 December 2019

Accepted: 27 February 2020

Published: 18 March 2020

Citation:

Wen X, Gao L and Hu Y (2020)
LAcModule: Identification of
Competing Endogenous RNA
Modules by Integrating Dynamic
Correlation. *Front. Genet.* 11:235.
doi: 10.3389/fgene.2020.00235

Competing endogenous RNAs (ceRNAs) regulate each other by competitively binding microRNAs they share. This is a vital post-transcriptional regulation mechanism and plays critical roles in physiological and pathological processes. Current computational methods for the identification of ceRNA pairs are mainly based on the correlation of the expression of ceRNA candidates and the number of shared microRNAs, without considering the sensitivity of the correlation to the expression levels of the shared microRNAs. To overcome this limitation, we introduced liquid association (LA), a dynamic correlation measure, which can evaluate the sensitivity of the correlation of ceRNAs to microRNAs, as an additional factor for the detection of ceRNAs. To this end, we firstly analyzed the effect of LA on detecting ceRNA pairs. Subsequently, we proposed an LA-based framework, termed LAcModule, to identify ceRNA modules by integrating the conventional Pearson correlation coefficient and dynamic correlation LA with multi-view non-negative matrix factorization. Using breast and liver cancer datasets, the experimental results demonstrated that LA is a useful measure in the detection of ceRNA pairs and modules. We found that the identified ceRNA modules play roles in cell adhesion, cell migration, and cell-cell communication. Furthermore, our results show that ceRNAs may represent potential drug targets and markers for the treatment and prognosis of cancer.

Keywords: ceRNA, microRNA, correlation, liquid association, modules

INTRODUCTION

MicroRNAs (length: ~22 nt) are a kind of small non-coding RNAs (Yates et al., 2013). They can interact with Argonaute protein to form the RNA-induced silencing complex. This complex binds to target RNA sequences (termed microRNA response elements, MRE) with partial complementarity, influencing the stability of target RNAs (Bartel, 2009; Yates et al., 2013). Recent studies revealed that different RNAs with microRNA response elements that bind to the same microRNAs can regulate each other by competitively binding to the microRNAs they share. This is known as the competing endogenous RNA (ceRNA) model, and these RNAs are termed ceRNAs (Salmena et al., 2011). CeRNAs can be messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), pseudogene gene transcripts, and circular RNAs (circRNAs). The detection of ceRNA can be used to explain the function of thousands of uncharacterized non-coding RNAs, and is also considered the “Rosetta stone of a hidden RNA language” (Salmena et al., 2011; Thomson and Dinger, 2016). In addition, ceRNAs play critical roles in post-transcriptional regulation. Thus,

they are involved in numerous physiological and pathological progresses, such as cancer (Salmena et al., 2011; Karreth and Pandolfi, 2013; Tay et al., 2014; Qi et al., 2015). For example, the tumor suppressor gene *PTEN* has been demonstrated to compete for microRNAs shared with other transcripts, such as *ZEB2*, *CNOT6L*, *VAPA*, *VCAN*, in many types of cancer (e.g., glioblastoma, melanoma, prostate cancer, and breast cancer) (Tay et al., 2014; Poliseno and Pandolfi, 2015). Of note, *PTENP1* (the pseudogene of *PTEN*), regulates the RNA levels of its cognate gene in prostate, melanoma, and clear-cell renal cell carcinoma (Poliseno et al., 2010, 2011; Johnsson et al., 2013; Yu et al., 2014). Furthermore, lncRNAs and circRNAs are also important regulators in the ceRNA model (Schmitt and Chang, 2016; Zhong et al., 2018). *HULC*, which is significantly upregulated in hepatocellular carcinoma, can reduce the expression and activity of miR-372 in liver cancer, which derepresses the translation of its target gene *PRKACB* and induces phosphorylation of the cAMP response element binding protein in liver cancer (Wang et al., 2010). CircHIPK3 inhibits the activity of miR-124 and promotes the expression of IL-6R by competitively binding to miR-124 (Zheng et al., 2016).

Owing to the large number of candidate ceRNA pairs and high cost of biological experiments, computational methods have become an efficient approach for the study of the ceRNA model (Le et al., 2017). For example, Zhou et al. (2014) constructed and investigated a ceRNA network in breast cancer; Sumazin et al. (2011) proposed a method based on conditional mutual information to infer candidate ceRNAs and analyze the ceRNA network in glioblastoma. Paci et al. (2014) used sensitivity correlation (SI) to infer the ceRNA network between lncRNAs and mRNAs in breast cancer. In that model, SI equals the difference between the Pearson correlation coefficient (PCC) and the partial correlation coefficient of ceRNAs with respect to their shared microRNAs. And List et al. (2019) further improved this method. All these methods predict ceRNA pairs based on two aspects: (1) ceRNAs should share a sufficient number of microRNAs, which can be evaluated through statistical tests, such as the hyper-geometric test (Salmena et al., 2011; Le et al., 2017); and (2) the expression of ceRNAs should be positively correlated, which can be estimated using the PCC (Chiu et al., 2015), SI (Paci et al., 2014; Do and Bozdogan, 2018), or conditional mutual information (Sumazin et al., 2011). In addition, Zhang et al. (2018) proposed LncmiRSRN to construct lncRNA-mRNA ceRNA network via estimating the causal effects of lncRNAs on mRNAs with the IDA method. Besides studying ceRNAs using sequencing data, researchers also proposed mathematical models to simulate the ceRNA process, such as the minimal model (Figliuzzi et al., 2013), mass-action model (Ala et al., 2013), and coarse-grained model (Yuan et al., 2015). Recently, Wei et al. constructed a unified coarse-grained competition motif model and uncovered the complexity and generality of the molecular competition effect, including the ceRNA model (Wei et al., 2019). In that study, the investigators proposed that the strength of competition between ceRNAs is influenced by the abundance of ceRNAs and their target microRNAs (Wei et al., 2019). The strength of the correlation is maximized in the “R near-equimolar” regime, and gradually decreases with concentration

of microRNAs away from the regime (Martirosyan et al., 2019; Wei et al., 2019). This means that the strength of ceRNA regulation varies based on the concentration of microRNAs. Those results also indicated that the strength of ceRNA regulation is sensitive to the expression levels of microRNAs around the proper concentration. Intuitively, two RNAs may be co-expressed due to some biological mechanisms. If they have an additional ceRNA relationship, their co-expression should be improved. When the levels of microRNA expression are low, the regulation between ceRNAs is not obvious. Furthermore, the ceRNA strength is sensitive to the expression levels of microRNAs. Hence, the strength of the co-expression should also be sensitive to these levels. We found the current studies using RNA sequencing data did not consider this characteristic of the ceRNA model. Hence, we propose that the sensitivity of co-expression to the expression levels of microRNAs they share may be a factor for predicting ceRNA relationships. We used dynamic correlation measure, termed liquid association (LA), to assess this sensitivity.

Unlike conventional correlations (e.g., PCC), dynamic correlation focuses on the change in the correlation of two variables following alterations in a third variable (Gunderson and Ho, 2014; Yu, 2018). For example, LA is defined as the mean of the derivative of the correlation between two objects with respect to a third condition (Li, 2002). LA has been used to identify disease candidate genes (Li et al., 2007) and human age-associated genes (Yang et al., 2018), as well as discover key microbial species and environmental factors of the microbial community (Ai et al., 2019).

LA is an appropriate measure for the evaluation of the correlation sensitivity of ceRNAs to microRNAs. In this study, we firstly analyzed the effectiveness of LA in detecting ceRNA pairs. Subsequently, we proposed a framework to investigate LA-based ceRNA modules (LAcModule) by integrating the conventional PCC and dynamic correlation LA with multi-view non-negative matrix factorization (NMF). By performing further analysis in breast cancer, we revealed that ceRNAs play roles in cell adhesion, cell migration, and cell-cell communication. Our results also showed that ceRNAs may represent promising drug targets and markers for the treatment and prognosis of cancer.

RESULTS

LA for the Prediction of ceRNA Pairs

Current studies often use the PCC or SI to detect ceRNA pairs. This approach ignores the sensitivity of the correlation between RNAs to the expression levels of their shared microRNAs. To overcome this limitation, we used LA (Li, 2002) to measure the dynamic change of the correlation for a ceRNA pair depending on the expression levels of their shared microRNAs. Suppose that EXP_{R1} and EXP_{R2} represent the expression levels of two ceRNA candidates $R1$ and $R2$, respectively, while EXP_{MIC} denotes the sum of the expression levels of all their shared microRNAs, MIC . We normalized EXP_{R1} and EXP_{R2} using the z-scoring method such that $E(EXP_{R1}) = E(EXP_{R2}) = 0$, $Var(EXP_{R1}) = Var(EXP_{R2}) = 1$, where $E(\cdot)$ and $Var(\cdot)$ represent the expectation and variance of a random variable, respectively.

Supposing the above, the PCC between $R1$ and $R2$ is:

$$\begin{aligned}\rho(R1, R2) &= \frac{E[(EXP_{R1} - E(EXP_{R1})) \times (EXP_{R2} - E(EXP_{R2}))]}{\sqrt{\text{Var}(EXP_{R1}) \times \text{Var}(EXP_{R2})}} \\ &= E(EXP_{R1} \times EXP_{R2}).\end{aligned}$$

By conditioning, $E(EXP_{R1} \times EXP_{R2}) = E(E(EXP_{R1} \times EXP_{R2} | EXP_{MIC})) = E(g(EXP_{MIC}))$, where $g(EXP_{MIC}) = E(EXP_{R1} \times EXP_{R2} | EXP_{MIC} = exp_{MIC})$ is the correlation between $R1$ and $R2$ when the expression level of the shared microRNA is exp_{MIC} .

The LA of $R1$ and $R2$ with respect to their shared microRNAs is defined as $LA(R1, R2 | MIC) = E(g'(EXP_{MIC}))$, where $g'(EXP_{MIC}) = E(EXP_{R1} \times EXP_{R2} | EXP_{MIC} = exp_{MIC})$. According to the Stein Lemma (Stein, 1981), if the sum of the expression levels of all the shared microRNAs MIC follows the standard normal distribution, $LA(R1, R2 | MIC) = E(EXP_{R1} \times EXP_{R2} \times EXP_{MIC})$, the calculation of LA can be simplified as shown below:

$$LA(R1, R2 | MIC) = \frac{\sum_{i=1}^N EXP_{R1_i} \times EXP_{R2_i} \times EXP_{MIC_i}}{N}$$

where N is the number of sample. We performed data transformation on EXP_{MIC} using the Van der Waerden's method to ensure that EXP_{MIC} follows the standard normal distribution. For $EXP_{MIC_1}, EXP_{MIC_2}, \dots, EXP_{MIC_N}$, we initially obtained their ranks r_1, r_2, \dots, r_N , and subsequently computed the transformed value as follows:

$$\begin{aligned}EXP_{MIC_1} &= \Phi^{-1}\left(\frac{r_1}{N+1}\right), EXP_{MIC_2} = \Phi^{-1}\left(\frac{r_2}{N+1}\right), \dots, \\ EXP_{MIC_N} &= \Phi^{-1}\left(\frac{r_N}{N+1}\right),\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

We downloaded the expression profiles of mRNAs, lncRNAs, and microRNAs of 1,072 and 365 patients with BRCA and LIHC, respectively, from The Cancer Genome Atlas (TCGA) to investigate the effect of LA on the prediction of ceRNA pairs. Non-expressed genes across all samples in a given type of cancer were removed (Figure 1A). Subsequently, we downloaded 2,667 experimentally validated ceRNA pairs from the LncCeRBase (Pian et al., 2018), miRSponge (Wang et al., 2015b), and lncACTdb (Wang et al., 2015a) databases. Considering the tissue-specific characteristics of ceRNAs and gene symbol mapping, seven validated ceRNA pairs in breast cancer and six validated ceRNA pairs in liver cancer were obtained (Table 1). All pairs in our candidate ceRNAs set significantly shared microRNAs (hypergeometric test, $p < 0.05$). For each disease, we considered these ceRNA pairs as benchmarks and all pairs, which were consisted of two genes in benchmarks and had sufficient number of shared microRNAs, as candidate pairs. We use these two datasets to evaluate the performance of LA-, PCC-, and SI-based methods, respectively. In breast cancer, the area under curve (AUC) values of the LA, PCC, and SI approaches were 0.58, 0.45, and 0.24, respectively. In liver cancer, these AUC values were 0.46, 0.27, and 0.35, respectively (Figure 2A). PCC and SI are usually

used to identify ceRNA pairs, and our results indicate that LA also has the ability to predict ceRNA pairs.

Identification of ceRNA Modules Using the LAcModule

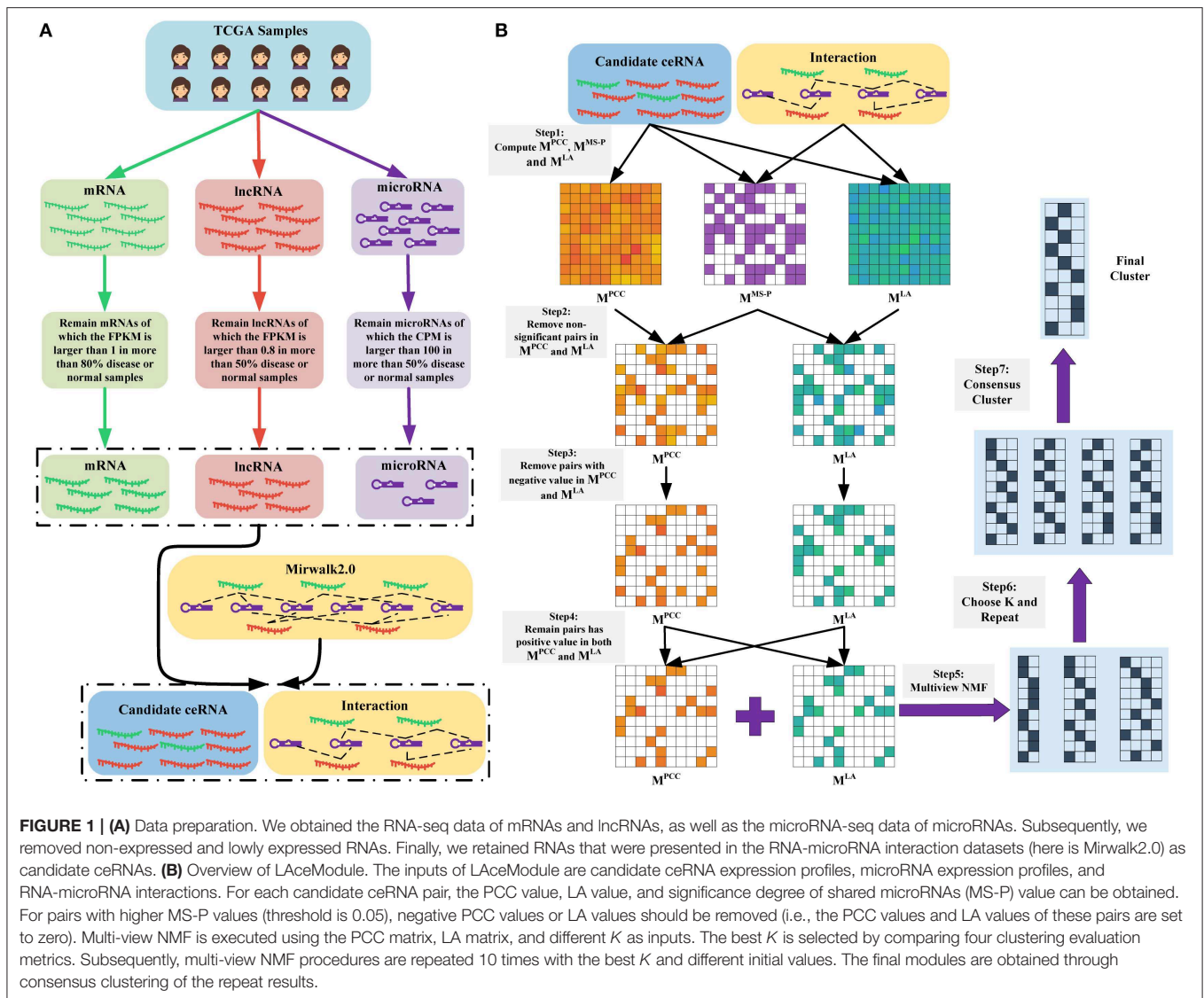
We proposed the LAcModule (Figure 1B), a framework based on multi-view NMF (Liu et al., 2013) to systematically identify ceRNA modules using LA. For each candidate ceRNA pair, we computed the PCC value, LA value, and the degree of significance of shared microRNAs (MS-P) (see section Materials and Methods) to construct three matrices M^{PCC} , M^{LA} , and M^{MS-P} , respectively. Subsequently, when the MS-P values of candidate ceRNA pairs were ≥ 0.05 , we set their corresponding PCC values and LA values to zero. Owing to the non-negativity requirement in the multi-view NMF framework, we set negative values in M^{PCC} and M^{LA} to zero. Considering that a ceRNA pair should be co-expressed and sensitive to change in the expression of their shared microRNAs, we set the values in the same entry of M^{PCC} and M^{LA} of candidate ceRNA pairs to zero if either of these values was zero. Finally, we integrated M^{PCC} and M^{LA} using multi-view NMF to identify ceRNA modules.

For multi-view NMF, there are two observation views $M = \{M^{PCC}, M^{LA}\}$, each of which is a $G \times G$ non-negative matrix, where G is the number of candidate ceRNAs. Each matrix in M , $M^v \in \{M^{PCC}, M^{LA}\}$, can be factorized to $U_{G \times K}^v \geq 0$ and $(V_{G \times K}^v)^T \geq 0$ that $M^v \approx U^v (V^v)^T$ and each row of $(V^v)^T$ can be considered as the K -rank representation of the corresponding candidate ceRNA point. Here, we attempted to identify a low-rank representation that is suitable for both views, which is defined as $(V^*)^T$. We factorized each matrix in M and made each $(V^v)^T$ as close as possible to $(V^*)^T$. Therefore, we determined the objective function as follows:

$$\begin{aligned}\min &\left(\|M^{PCC} - U^{PCC} (V^{PCC})^T\|_F^2 + \|M^{LA} - U^{LA} (V^{LA})^T\|_F^2 \right. \\ &\left. + \lambda_{PCC} \|V^{PCC} - V^*\|_F^2 + \lambda_{LA} \|V^{LA} - V^*\|_F^2 \right) \\ \text{s.t. } &\forall 1 \leq k \leq K, \|U_{:,k}^{PCC}\|_1 = 1, \|U_{:,k}^{LA}\|_1 \\ &= 1 \text{ and } U^{PCC}, U^{LA}, V^{PCC}, V^{LA}, V^* \geq 0\end{aligned}$$

where λ_{PCC} and λ_{LA} tunes the relative weight among different views and between standard NMF error and disagreement among $(V^*)^T$, $(V^{PCC})^T$, and $(V^{LA})^T$. We used an iterative procedure by updating one variable, while maintaining the remaining variables fixed to solve this optimization problem (see details in Materials and Methods section). After computing the $(V^*)^T$, we obtained the module label of RNA i using $\text{argmax}_{j=1,2,\dots,K} V_{ij}^*$.

Of note, the LAcModule requires pre-determination of the number of modules, K . We evaluated the clustering performance to select an optimal K ranging from 10 to 400 with an increment of 10 by considering four metrics (Figures 2B,C), namely the C-index (Hubert and Schultz, 1976), McClain-Rao (McClain and Rao, 1975), point biserial correlation coefficient (Milligan, 1981), and silhouette coefficient (Rousseeuw, 1987). By simultaneously considering four metrics on two matrices, we selected $K = 360$ in BRCA and $K = 370$ in LIHC. To obtain robust ceRNA modules,



the LAcModule repeated the multi-view NMF procedures 10 times, and computed a consensus matrix to identify ceRNA modules using the cluster-based similarity partitioning algorithm (CSPA) (Strehl and Ghosh, 2003). Specifically, CSPA generates a binary matrix for each result of the multi-view NMF clustering, with “1” representing two associated genes in the same cluster and “0” for not. The consensus matrix is the sum of these binary matrices. ceRNA modules can be identified through spectral clustering on this consensus matrix using the optimal K selected above.

Comparison Between the LAcModule and PCC/SI-Based Methods

We used NMF to replace the multi-view NMF and the PCC matrix or SI matrix as input to compare the performance of conventional and dynamic correlations in the detection of ceRNA modules. In the PCC matrix and SI matrix, negative values or the corresponding MS-P values ≥ 0.05 were set to zero. We also tested

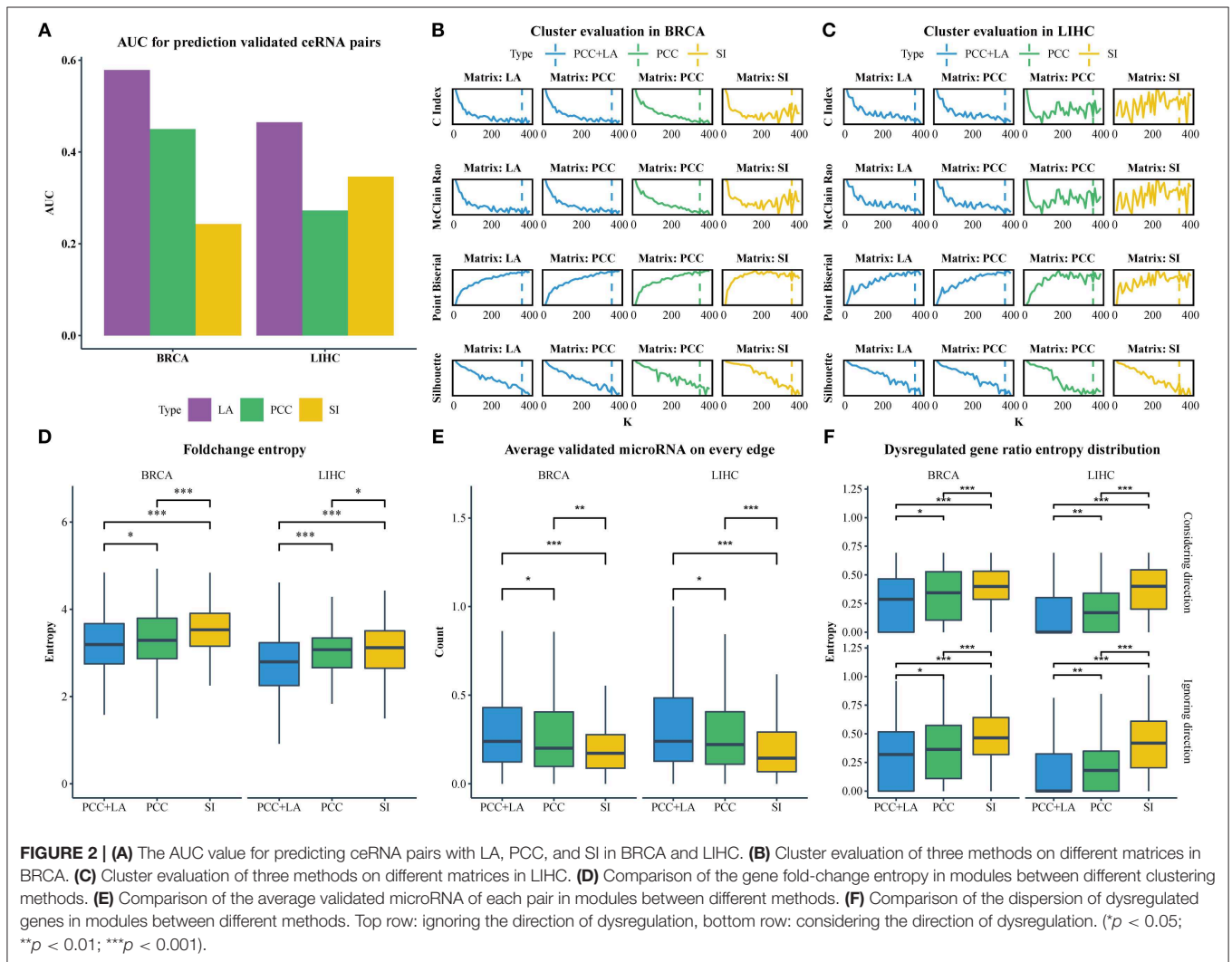
K ranging 10–400, with an increment of 10, and evaluated the clustering performance with the same indicators mentioned in Section Identification of ceRNA modules using the LAcModule. We selected K s equal to 350 and 360 for PCC-based and SI-based results in BRCA, respectively, while K s equal to 360 and 340, respectively, were selected for LIHC (Figures 2B,C). In the following sections, we used “PCC+LA” to represent the modules detected by the LAcModule, as well as “PCC” and “SI” to represent the modules based on PCC or SI, respectively.

CeRNA pairs are highly co-expressed; hence, the fold change of gene expression in a ceRNA module tends to be similar in the disease and normal states. We analyzed the differential expression of genes between normal tissues and tumor tissues using the R packages edgeR (Robinson et al., 2010) (Materials and Methods section) to obtain the fold change. Subsequently, we calculated the entropy of fold change for each module. As shown in Figure 2D, the entropy distribution of “PCC+LA” was significantly lower than those of “PCC” (FDR = 0.0476 in BRCA,

TABLE 1 | LA, PCC, and SI values of validated ceRNA pairs.

ceRNA1	ceRNA2	PCC	LA	SI	MS-P*	Disease
ENSG00000234741	ENSG00000171862	-0.058	0.040	-0.008	0.005	BRCA
ENSG00000251562	ENSG00000070831	0.043	-0.009	0.002	0.001	BRCA
ENSG00000251562	ENSG00000135446	-0.377	0.000	-0.003	0.022	BRCA
ENSG00000115414	ENSG00000026508	0.082	-0.003	-0.001	0.001	BRCA
ENSG00000108821	ENSG00000026508	-0.014	0.082	0.001	0.029	BRCA
ENSG00000171862	ENSG00000038427	0.379	0.075	-0.004	0.002	BRCA
ENSG00000038427	ENSG00000139687	0.368	0.058	0.000	0.003	BRCA
ENSG00000226950	ENSG00000168036	0.131	0.103	-0.003	0.012	LIHC
ENSG00000234741	ENSG00000150593	0.205	-0.205	-0.014	0.003	LIHC
ENSG00000234741	ENSG00000171862	-0.003	-0.107	-0.002	0.013	LIHC
ENSG00000241388	ENSG00000057663	0.035	-0.068	-0.005	0.033	LIHC
ENSG00000251164	ENSG00000148516	-0.093	0.097	-0.001	0.004	LIHC
ENSG00000251164	ENSG00000168615	-0.392	0.411	0.003	0.034	LIHC

*MS-P, statistical significance of shared microRNAs between RNAs.



FDR = 5.12E-06 in LIHC; Kolmogorov–Smirnov one-tailed test) and “SI” (FDR = 8.16E-09 in BRCA, FDR = 5.12E-06 in LIHC; Kolmogorov–Smirnov one-tailed test). By comparing “PCC” and “SI,” we also found that the former was significantly lower than the latter (FDR = 7.20E-05 in BRCA, FDR = 4.76E-02 in LIHC; Kolmogorov–Smirnov one-tailed test). Similarly, if a gene in a ceRNA pair is dysregulated, the other gene also tends to be dysregulated. In addition, the direction of the dysregulation also tends to be the same. This indicates that, when a gene in a ceRNA pair is upregulated, the other gene also tends to be upregulated. Therefore, we calculated the entropy of the dysregulated gene ratio and the entropy of the different dysregulation direction ratio of each module. The results are shown in **Figure 2F**; the top row illustrates the dysregulated gene ratio without direction, while the bottom row shows the dysregulated gene ratio with direction. The results showed that the modules of “PCC+LA” were significantly lower than those of “PCC” (top: FDR = 0.0246 in BRCA, FDR = 6.50E-03 in LIHC; bottom: FDR = 2.62E-02 in BRCA, FDR = 7.82E-03 in LIHC; Wilcoxon one-tailed test) and “SI” (top: FDR = 5.82E-12 in BRCA, FDR = 3.47E-24 in LIHC; bottom: FDR = 6.88E-10 in BRCA, FDR = 2.59E-23 in LIHC; Wilcoxon one-tailed test) in both situations. For “PCC” and “SI,” the former performed better than the latter in two situations (top: FDR = 5.39E-06 in BRCA, FDR = 9.15E-16 in LIHC; bottom: FDR = 2.39E-04 in BRCA, FDR = 2.09E-15 in LIHC; Wilcoxon one-tailed test).

CeRNAs are regulated through shared microRNAs. Therefore, ceRNA modules may tend to share more microRNAs in each pair. We used experimentally validated mRNA-microRNA interaction in miRTarBase (Chou et al., 2016) to evaluate the average number of shared microRNAs in a pair. The results are shown in **Figure 2E**. The modules of “PCC+LA” shared more microRNAs on average than those of “PCC” (FDR = 1.84E-02 in BRCA, FDR = 1.84E-02 in LIHC; Wilcoxon one-tailed test) and “SI” (FDR = 1.05E-06 in BRCA, FDR = 2.62E-09 in LIHC; Wilcoxon one-tailed test). Moreover, the modules of “PCC” shared more microRNAs on average than those of “SI” (FDR = 8.46E-03 in BRCA, FDR = 3.82E-05 in LIHC; Wilcoxon one-tailed test).

Collectively, the comparisons of gene fold change, gene dysregulation ratio, and number of shared microRNAs suggest that the integration of conventional and dynamic correlations offers better detection of ceRNA modules than conventional correlation alone.

Functional Analysis of ceRNA Modules in Breast Cancer

We used the results of the BRCA dataset for further analysis. For investigating the difference in ceRNA relationship between tumor and normal states, we identified 348 ceRNA modules in breast tumor tissues and 314 modules in normal breast tissues using LAcModule. We used the g:Profiler (Reimand et al., 2007) for function enrichment of ceRNA modules in the disease and normal states. In **Table 2**, we listed the top five most frequently enriched Gene Ontology (GO) terms in both states. We analyzed the relationship of these GO terms (**Figure 3A**) and found that the ceRNA modules are associated with cell

TABLE 2 | Most frequently enriched GO terms.

Term ID	Domain	Term name	Count
Disease state			
GO:0007155 ¹	BP	Cell adhesion	15
GO:0022610 ¹	BP	Biological adhesion	15
GO:0032501 ¹	BP	Multicellular organismal process	13
GO:0006955 ¹	BP	Immune response	11
GO:0007165 ³	BP	Signal transduction	11
GO:0009605 ¹	BP	Response to external stimulus	11
GO:0032502 ¹	BP	Developmental process	11
Normal state			
GO:0010033 ²	BP	Response to organic substance	8
GO:0070887 ²	BP	Cellular response to chemical stimulus	8
GO:0071310 ²	BP	Cellular response to organic substance	8
GO:0007154 ²	BP	Cell communication	7
GO:0007165 ³	BP	Signal transduction	7
GO:0007166 ²	BP	Cell surface receptor signaling pathway	7
GO:0023052 ²	BP	Signaling	7
GO:0042221 ²	BP	Response to chemical	7
GO:0050896 ²	BP	Response to stimulus	7
GO:0051716 ²	BP	Cellular response to stimulus	7

¹Disease state-specific GO terms, ²normal state-specific GO terms, and ³common GO terms in both disease state and normal state.

adhesion (GO:0007155) specifically in disease tissues, compared with the modules of the normal tissues. It is established that activation of invasion and metastasis is an important hallmark of cancer (Hanahan and Weinberg, 2011). Current research studies suggest that the loss of cell adhesion is strongly associated with tumor invasion and metastasis (Cavallaro and Christofori, 2001; Okegawa et al., 2004). This term and its parent term biological adhesion (GO:0022610) are enriched in 15 modules. In most of these modules, gene pairs have larger PCC and LA values in diseased cells than in normal cells (**Figure 3B**, **Supplementary File 1**). Similarly, the most frequently and specifically enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (**Supplementary Table 1**) in the disease state were cell adhesion molecules (CAMs) (KEGG:004514) and focal adhesion (KEGG:04510), which play a pivotal role in tumor metastasis and cell migration (Ridley et al., 2003; Okegawa et al., 2004; Kim and Wirtz, 2013). In most of these modules, gene pairs also have larger PCC and LA values in diseased cells than in normal cells (**Figures 3C,D**, **Supplementary Files 2, 3**). These results suggest that ceRNAs play important roles in the invasion and metastasis of breast tumor cells, as previously indicated (Yang et al., 2014; Hu et al., 2017; Zheng et al., 2018).

As shown in **Table 3**, we also investigated the most significant terms. The most significant GO terms were obtained from Module 199, and were associated with defense against other organisms, such as viruses. Numerous studies have indicated that the mouse mammary tumor virus, bovine leukemia virus, human papillomavirus, and Epstein–Barr virus are associated with breast cancer (Amarante and Watanabe, 2009; Alibek et al., 2013; Lawson et al., 2018). As shown in **Figure 4A**, genes in this

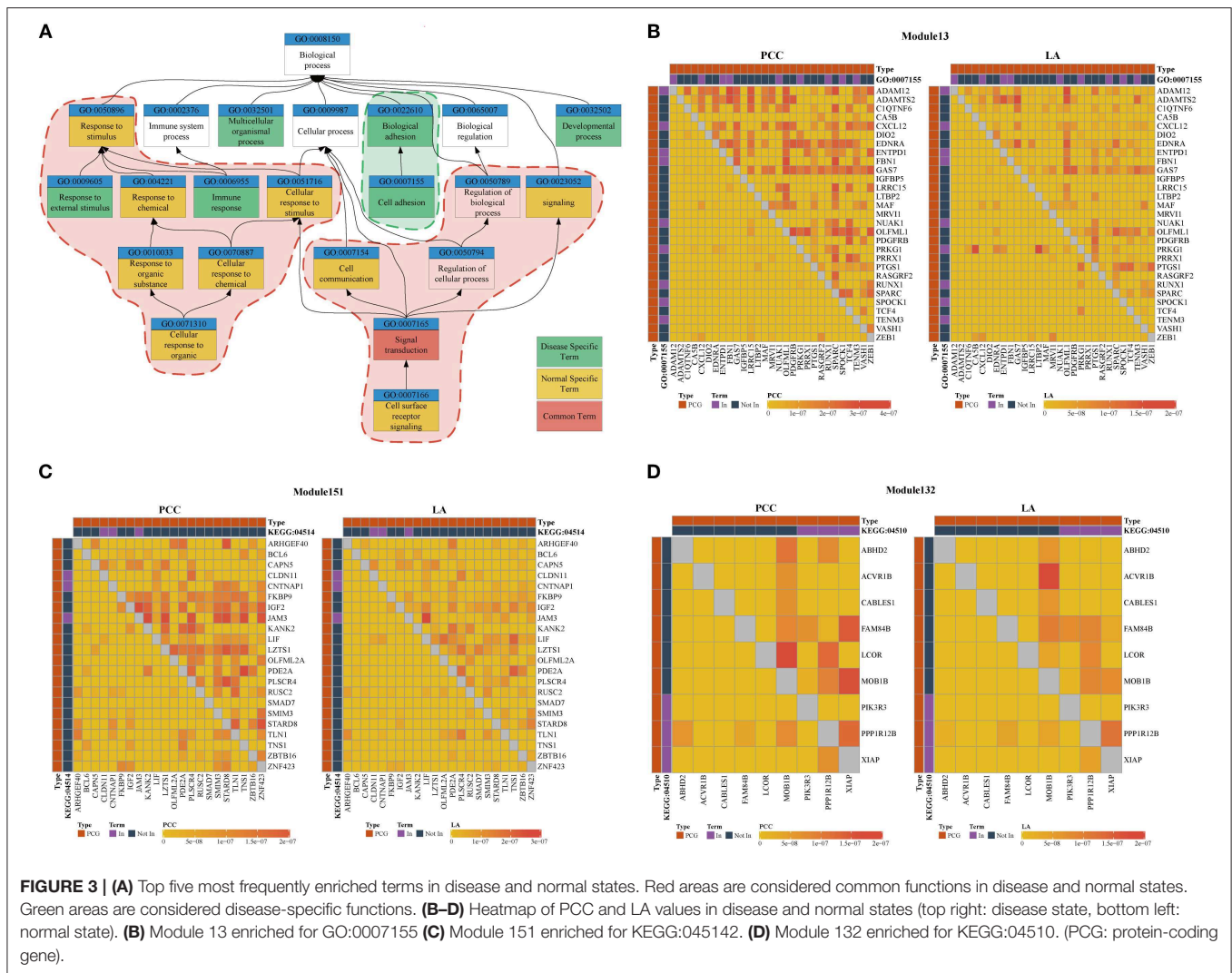


FIGURE 3 | (A) Top five most frequently enriched terms in disease and normal states. Red areas are considered common functions in disease and normal states. Green areas are considered disease-specific functions. **(B–D)** Heatmap of PCC and LA values in disease and normal states (top right: disease state, bottom left: normal state). **(B)** Module 13 enriched for GO:0007155 **(C)** Module 151 enriched for KEGG:045142. **(D)** Module 132 enriched for KEGG:04510. (PCG: protein-coding gene).

module exhibit larger PCC and LA values. For the KEGG dataset, the most significant pathways were obtained from Module 103 (Figure 4B); these pathways were significantly enriched for oxidative phosphorylation, thermogenesis, Huntington’s disease, and Parkinson’s disease.

Collectively, comparison of the functions of ceRNA modules in breast tumor and normal tissues suggested that the ceRNA relationship may rewrite in different cell states and ceRNA may exert an effect on the development or progression of breast cancer.

ceRNA Modules Are Associated With Aberrant Genetic and Epigenetic Regions in Breast Cancer

According to the methods described in Materials and Methods section, we obtained 1,829 dysregulated mRNAs, 264 dysregulated lncRNAs (Supplementary File 4, Supplementary Table 2), 1,074 CNV genes, and 51 differentially methylated genes. We considered these genes as BRCA-associated genes. We identified BRCA-associated modules that

are enriched for both expression-dysregulated genes ($p < 0.01$); and genes associated with aberrant CNV or DNA methylation ($p < 0.05$). Totally, we obtained eight BRCA associated modules (Table 4, Figure 5A, Supplementary File 5). The top five most significant enriched functions and pathways of these modules are shown in Figures 5B,C. We found that six of eight modules are significantly enriched for breast cancer-associated functions and pathways, such as immune response (Module 90 and Module 309) (Bates et al., 2018), cell communication (Module 110) (Brooks and Wicha, 2015), cell cycle (Module 73) (Otto and Sicinski, 2017), blood vessel morphogenesis (Module 45) (Kakolyris et al., 2000), and sucrose process (Module 37) (Jiang et al., 2016).

We also investigated the relationship between lncRNAs and oncogenes in these modules. We downloaded the oncogene dataset from ONGENE (Liu et al., 2017). We found that in Module 45, *LINC01485* was correlated with *AQP1*, with PCC and LA values of 0.397 and 0.087, respectively. *AQP1* is related to tumor cell migration, invasion, and angiogenesis (Tomita et al., 2017). In Module 309, lncRNA *HSPC324* (ENSG00000228401) was highly co-expressed with oncogene *KLF2*, with PCC and

TABLE 3 | Most significant GO terms and KEGG pathways.

Term ID	Domain	Term name	FDR	Precision	Module ID
GO:0051607	BP	Defense response to virus	3.81E-19	0.459	Module 199
GO:0009615	BP	Response to virus	1.70E-18	0.486	Module 199
GO:0043207	BP	Response to external biotic stimulus	8.96E-18	0.595	Module 199
GO:0051707	BP	Response to other organism	8.96E-18	0.595	Module 199
GO:0098542	BP	Defense response to other organism	8.96E-18	0.486	Module 199
KEGG:00190	KEGG	Oxidative phosphorylation	3.07E-09	0.217	Module 103
KEGG:03010	KEGG	Ribosome	2.30E-08	0.2	Module 228
KEGG:04714	KEGG	Thermogenesis	2.33E-08	0.246	Module 103
KEGG:05016	KEGG	Huntington's disease	2.33E-08	0.232	Module 103
KEGG:05012	KEGG	Parkinson's disease	2.33E-08	0.203	Module 103

LA values of 0.642 and 0.012, respectively. *KLF2* is a tumor suppressor gene, which inhibits cell growth, migration, and invasion in numerous types of cancer, such as colorectal cancer (Wang et al., 2017), breast cancer (Zhang et al., 2015), and prostate cancer (Wang et al., 2019). Our results indicate that these lncRNAs may be new drug targets for cancer therapy.

ceRNA Modules Predict Survival in Patients With Breast Cancer

MicroRNAs are clinically important in cancer. Therefore, we used the top 15 most commonly shared microRNAs in BRCA-associated modules to analyze their relationship with patient outcome. We used a k-means algorithm to classify patients into two groups based on the expression of the top 15 microRNAs in each module, and performed a Kaplan–Meier analysis. We found that half of the top 15 microRNAs sets in Module 45, Module 110, Module 90, and Module 309, could significantly distinguish patients (Figure 6A). By analyzing the microRNAs in these modules, we found that 10 of 30 microRNAs were shared by more than three modules (Figure 6B, Supplementary Table 3). In addition, half of those 10 microRNAs were derived from the *let-7* microRNA family. Many studies have suggested that the *let-7* family participates in the process of metastasis and resistance to therapy in breast cancer (Cunningham et al., 2010; Chiu et al., 2014). Our results also demonstrate that the *let-7* microRNA family can be a prognostic marker of breast cancer (Figure 6B, Supplementary File 9).

In addition, we identified five ceRNA pairs (i.e., *CDH5* with *FAM212B*, *CDH5* with *GYG2* in Module 45, *TRIB1* with *IL33*, *TRIB1* with *INMT*, and *TRIB1* with *MMRN1* in Module 110) that can be used as prognostic markers of breast cancer in BRCA-associated modules (Figures 6C,D, Supplementary File 9). In our dataset, *CDH5* is a hyper-methylated gene (six of eight sites are hypermethylated in the promoter region), while *TRIB1* is duplicated or amplified in >14% of patients with BRCA. Recent studies showed that *CDH5* levels and *CDH5* glycosylation are biomarkers for metastatic breast cancer (Fry et al., 2016) and *TRIB1* plays a critical role in cell cycle and survival via NF- κ B signaling (Gendelman et al., 2017). Interestingly, we found that these genes cannot individually act as prognostic markers (Supplementary File 6). However, the gene sets of the ceRNA

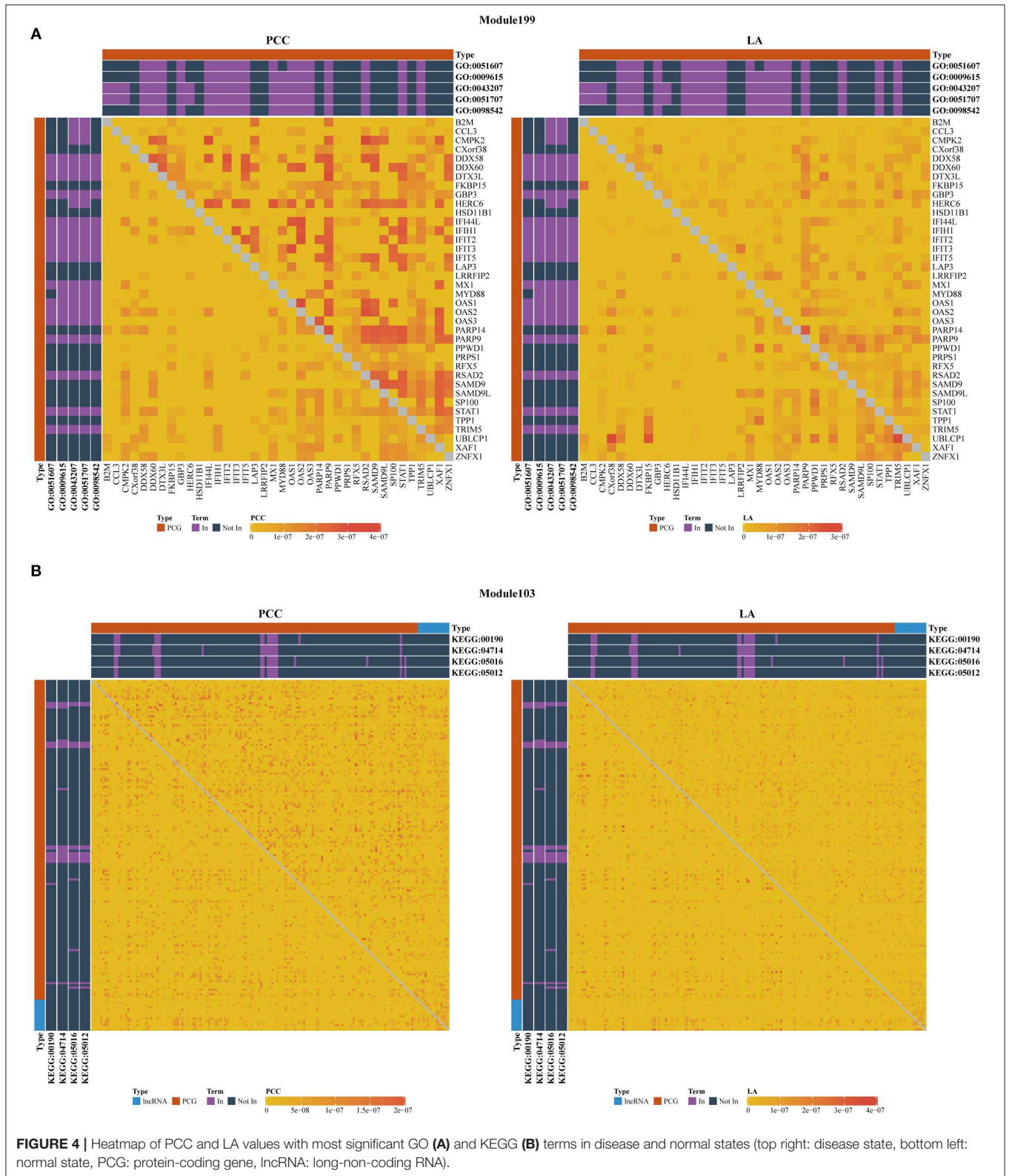
pairs and their experimental validated shared microRNAs in miRTarBase are effective markers for therapy and prognosis, indicating that these ceRNAs may collaborate in breast cancer.

Furthermore, we also identified ceRNA modules that can be considered prognostic markers for breast cancer. Similarly, we used the expression of ceRNAs in each module to classify patients into two groups via a k-means algorithm, and performed a Kaplan–Meier analysis. In total, we found six modules that can distinguish patients into two subgroups with significantly different survival times (log-rank test, $p < 0.01$). As shown in Figure 7 and Supplementary File 7, the patients with lower expressions in Module 63 and Module 270, as well as those with higher expression in Module 11, Module 25, Module 56, and Module 204 had longer survival time. Notably, Module 11 consists of lncRNAs. Collectively, these results confirm that ceRNAs play important roles in the treatment and prognosis of BRCA.

DISCUSSION

CeRNAs play critical roles in post-transcriptional regulation and are thus involved in many physiological and pathological progresses. An increasing number of non-coding RNAs that are able to influence their ceRNA partners via the ceRNA mechanism have been detected. Computational methods are an efficient approach for the detection and analysis of ceRNA relationships. By integrating the basic hypothesis and latest studies of ceRNA, we introduced dynamic correlation LA as one of the factors for detecting ceRNA pairs. Moreover, we integrated the multi-view NMF method with conventional PCC to detect ceRNA modules. Our results indicated that LA is effective in detecting ceRNA pairs and modules. The results of subsequent analysis showed that ceRNAs play important roles in breast cancer, especially in cell adhesion, cell migration, and cell-cell communication. Our results also demonstrated that ceRNAs may represent promising drug targets and markers for the treatment and prognosis of cancers.

In multi-view NMF, the parameters λ s balances the weight of different views. We set $\lambda_{COR} = 1$, and test $\lambda_{LA} \in \{1, 2, 3, 4, 5, 10, 15, 20, 25\}$ in the BRCA dataset. The results showed that there is no significant difference among these



parameters (Supplementary File 8), indicating that our method is not sensitive to these parameters.

The simplified calculation of LA demands that the distribution of the third variable follows standard normal distribution. We

used the Van der Waerden's method to transform the microRNA expression profile to ensure it follows the standard normal distribution. We tested the performance of this method. We got a number of random data, which followed the negative binomial

TABLE 4 | BRCA-associated ceRNA modules.

Module ID	Differentially expressed gene p -value	Methylation gene p -value	CNV gene p -value	LncRNA count
Module 37	1.26E-5***	0.20	4.7E-2*	2
Module 45	0***	1.3E-2*	0.89	1
Module 56	0***	9.7E-3**	0.62	6
Module 73	8.86E-7***	1	2.6E-2*	0
Module 90	0***	2.2E-3**	0.81	2
Module 110	0***	0.16	3.3E-3**	0
Module 128	2.24E-6*	2.1E-2*	0.51	0
Module 309	0***	2.1E-2*	0.96	7

Level of significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

distribution. The sizes of the data set ranged from 10 to 2,000, with an increment of 20. We transform them by using Van der Waerden's method, respectively. Then we got another random data set following the standard normal distribution, whose size ranged from 1,000 to 100,000 with an increment of 1,000. We tested if the distribution of each pair between these two sets was similar by using Kolmogorov–Smirnov test. The result was shown in **Supplementary File 10**. The result suggested that the distribution of each pair has no significant difference.

LA is very convenient for the measurement of dynamic correlation. However, according to the results reported by Wei (Wei et al., 2019), the ceRNA strength forms a sharp peak around the most optimal microRNA concentration, indicating very large deviation of correlation. In this study, we used LA as a characteristic of ceRNA regulation. When the data contain the microRNA concentration ranging from the lower situation to higher situation, or symmetrically distributed on both sides of the most optimal microRNA concentration, the value of LA may be close to zero (according to the definition of LA). A solution to this problem is to measure the mean of the absolute value of the deviation, $LA'(R1, R2|MIC) = E(|g'(EXP_{MIC})|)$, instead of $LA(R1, R2|MIC) = E(g'(EXP_{MIC}))$, where $g'(EXP_{MIC})$ is the deviation of correlation corresponding to EXP_{MIC} . Moreover, LA is developed based on linear correlation PCC. However, real systems are more likely non-linear. Additional studies are warranted to determine the method for the calculation of the dynamic correlation of non-linear correlation.

Furthermore, besides the positive correlation and dynamic correlation, ceRNA pairs may have additional unknown features, which are also worthy of further investigation. The multi-view NMF framework has potential to integrate all these features for more accurate identification of ceRNA modules (Wang et al., 2018).

MATERIALS AND METHODS

Datasets

We downloaded the RNA-seq data of mRNAs and lncRNAs, as well as the microRNA-seq data of patients with breast invasive carcinoma (BRCA) and liver hepatocellular carcinoma (LIHC)

from TCGA (Weinstein et al., 2013). For patients with BRCA, we obtained the fragments per kilobase of exon model per million reads mapped (FPKM) values and read counts of mRNAs and lncRNAs from 1,090 tumor samples and 113 normal samples; we also obtained the read counts of mature microRNAs from 1,077 tumor samples and 104 normal samples. For patients with LIHC, we obtained the FPKM values and read counts of mRNAs and lncRNAs from 370 tumor samples and 50 normal samples; we also obtained the read counts of mature microRNAs from 371 tumor samples and 50 normal samples. Both microRNA-mRNA and microRNA-lncRNA interactions were downloaded from the mirwalk2.0 database (Dweep and Gretz, 2015), which incorporates predicted interactions curated in at least two of 13 different RNA-microRNA interaction databases, such as DIANA (Maragkakis et al., 2011) and miRDB (Wang and El Naqa, 2008). For further analysis, we downloaded the copy number variation data and DNA methylation data of patients with BRCA from TCGA.

Data Preparation

Figure 1A shows the process of data preparation. Firstly, we only included samples that have complete mRNA, lncRNA, and microRNA expression data. We obtained 1,072 disease samples and 113 normal samples of BRCA, as well as 365 disease samples and 50 normal samples of LIHC. Subsequently, we excluded lowly expressed mRNAs, lncRNAs, and microRNAs. We only retained mRNAs with FPKM values >1 in $>80\%$ of disease or normal samples, lncRNAs with FPKM values >0.8 in $>50\%$ of disease or normal samples, and microRNAs with counts per million mapped reads values >100 in $>50\%$ of disease or normal samples (Mullokanov et al., 2012). Notably, we also excluded RNAs without interactions curated in the mirwalk2.0 database. Finally, we obtained 1,128 lncRNAs, 1,117 mRNAs, and 137 mature microRNAs in BRCA, as well as 621 lncRNAs, 8,783 mRNAs, and 138 mature microRNAs in LIHC. All retained mRNAs and lncRNAs were considered candidate ceRNAs. We transformed the expression profiles of the candidate ceRNAs through $\log(FPKM + 1)$ and those of mature microRNAs through $\log(CPM + 1)$.

Conventional Correlation Between RNAs

Current methods use conventional correlations as factors to predict ceRNA pairs. PCC and SI are most commonly used. We calculated the PCC value of a candidate ceRNA pair (e.g., $R1, R2$) as follows:

$$\rho(R1, R2) = \frac{E[(EXP_{R1} - E(EXP_{R1})) \times (EXP_{R2} - E(EXP_{R2}))]}{\sqrt{\text{Var}(EXP_{R1}) \times \text{Var}(EXP_{R2})}}$$

where EXP_{R1} and EXP_{R2} represent the paired expression of $R1$ and $R2$, respectively. For the SI value of $R1$ and $R2$, we initially added the expression of shared microRNAs between $R1$ and $R2$. Subsequently, we calculated the partial correlation with respect to their shared microRNAs as follows:

$$\rho(R1, R2|MIC) = \frac{\rho(R1, R2) - \rho(R1, MIC) \times \rho(R2, MIC)}{\sqrt{1 - \rho^2(R1, MIC)} \times \sqrt{1 - \rho^2(R2, MIC)}}$$

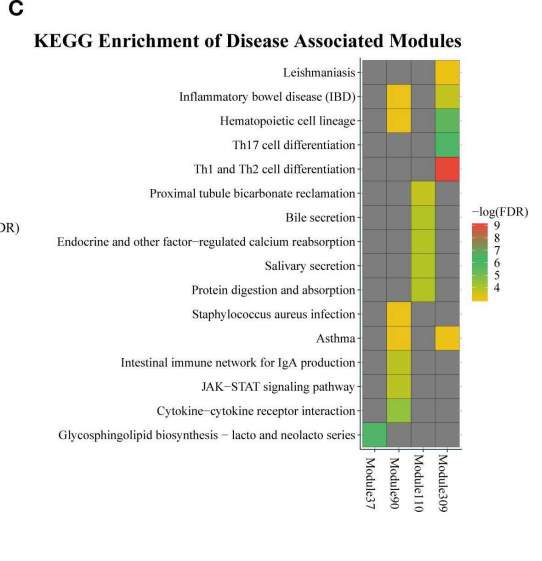
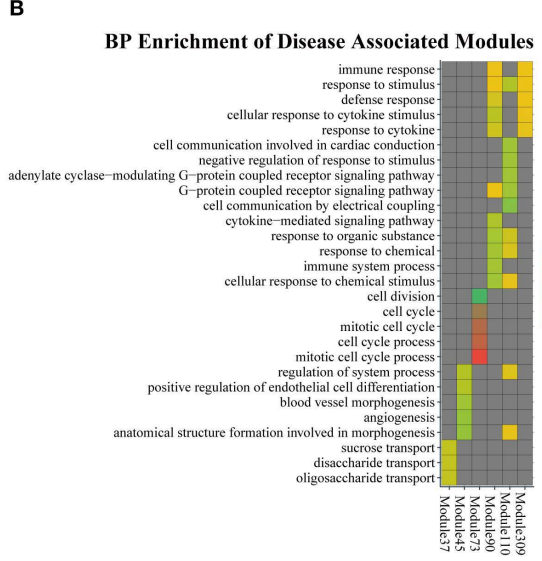
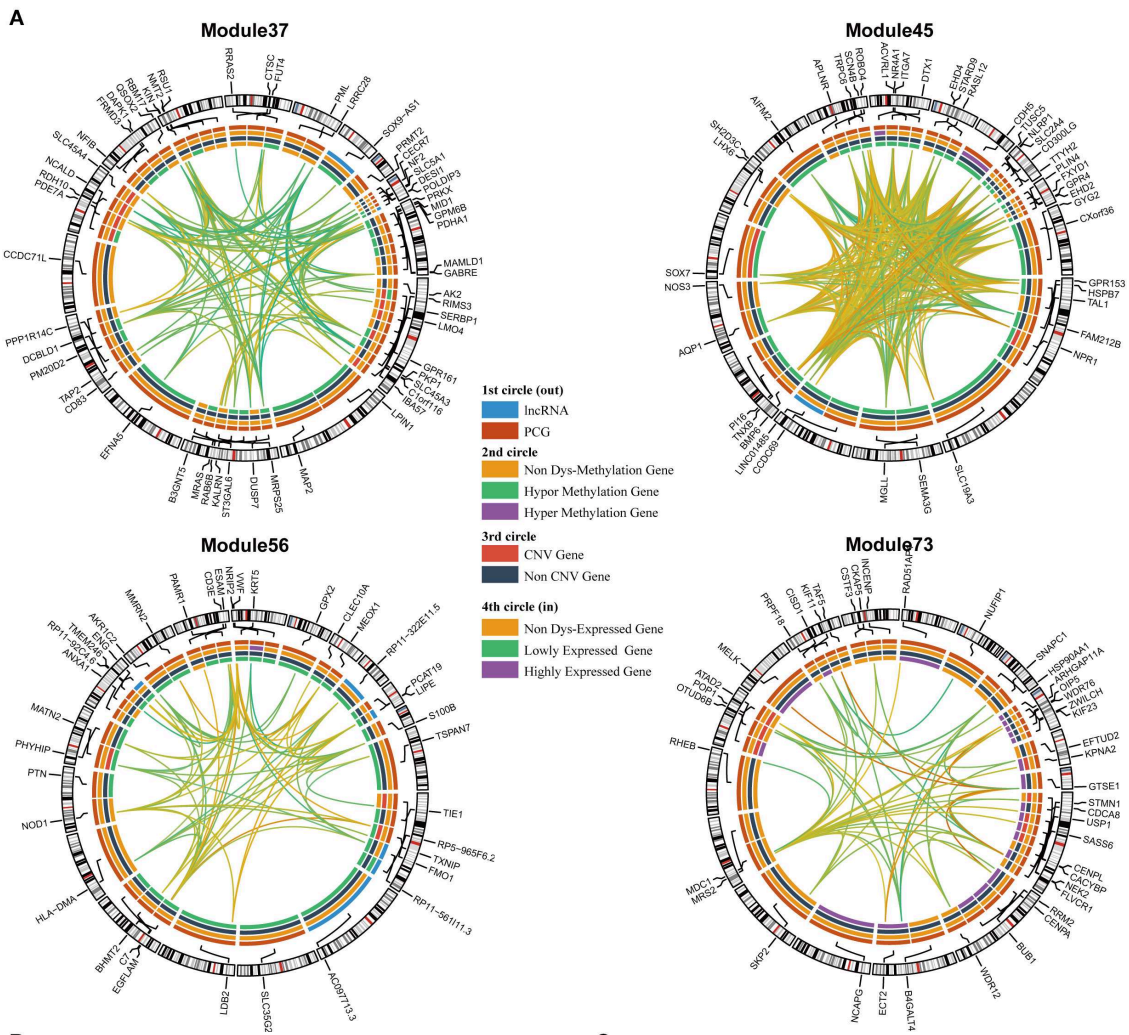


FIGURE 5 | (A) The modules that are enriched for breast invasive carcinoma (BRCA)-associated genes. **(B)** GO enriched terms of BRCA-associated modules. **(C)** KEGG enriched terms of BRCA-associated modules.

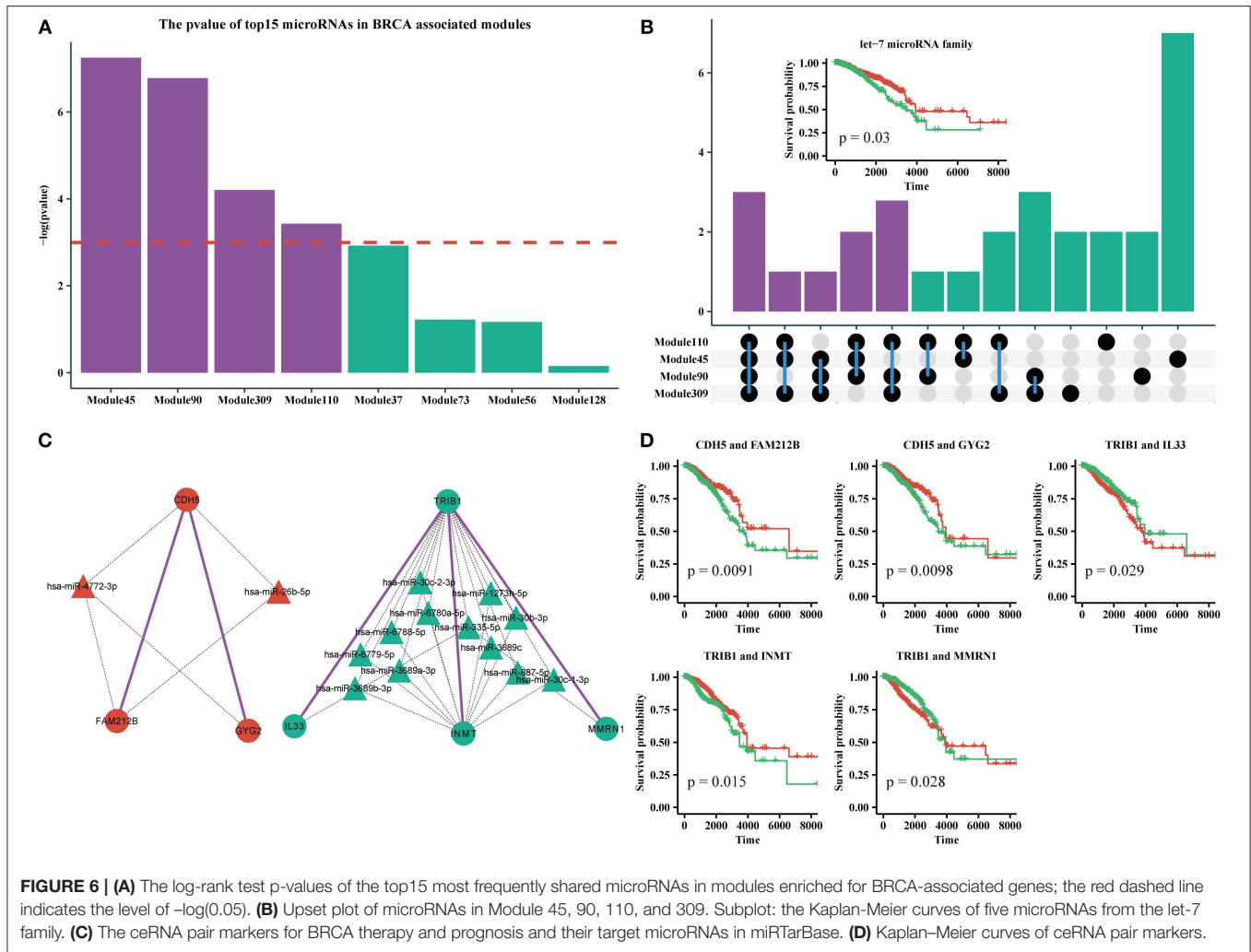


FIGURE 6 | (A) The log-rank test p-values of the top15 most frequently shared microRNAs in modules enriched for BRCA-associated genes; the red dashed line indicates the level of $-\log(0.05)$. **(B)** Upset plot of microRNAs in Module 45, 90, 110, and 309. Subplot: the Kaplan-Meier curves of five microRNAs from the let-7 family. **(C)** The ceRNA pair markers for BRCA therapy and prognosis and their target microRNAs in miRTarBase. **(D)** Kaplan-Meier curves of ceRNA pair markers.

where MIC indicates the shared microRNAs between $R1$ and $R2$. Finally, we computed the SI value as follows:

$$SI(R1, R2) = \rho(R1, R2) - \rho(R1, R2|MIC)$$

Statistical Significance of Shared MicroRNAs Between RNAs

CeRNAs are thought to usually share numerous microRNA targets. We used the hypergeometric test to calculate the significance of shared microRNAs for a given ceRNA pair (e.g., $R1$ and $R2$). The significance p -value can be obtained as follows:

$$p(R1, R2) = 1 - \sum_{i=0}^{q-1} \frac{\binom{T}{i} \times \binom{Q-T}{b-i}}{\binom{Q}{b}}$$

where Q is the total number of considered microRNAs, T is the number of microRNAs targeting candidate $R1$, b is the number of microRNAs targeting candidate $R2$, and q is the number of microRNAs targeting both candidates $R1$ and $R2$.

Solution of the Objective Function

According to the multi-view NMF framework, we obtained the objective function as follows:

$$\begin{aligned} \min & \left(\|M^{PCC} - U^{PCC}(V^{PCC})^T\|_F^2 + \|M^{LA} - U^{LA}(V^{LA})^T\|_F^2 \right. \\ & \left. + \lambda_{PCC} \|V^{PCC} - V^*\|_F^2 + \lambda_{LA} \|V^{LA} - V^*\|_F^2 \right) \\ \text{s.t. } & \forall 1 \leq k \leq K, \|U^{PCC}_{\cdot,k}\|_1 = 1, \|U^{LA}_{\cdot,k}\|_1 \\ & = 1 \text{ and } U^{PCC}, U^{LA}, V^{PCC}, V^{LA}, V^* \geq 0 \end{aligned}$$

where $\lambda_{PCC}, \lambda_{LA}$ tunes the relative weight among different views and between the standard NMF error and disagreement among $(V^*)^T, (V^{PCC})^T$ and $(V^{LA})^T$. We used an iterative update procedure by updating one variable and maintaining the remaining variables fixed to solve this optimization problem. The specific iteration rules are listed as follows:

1. fixing V^*, V^{PCC} , and V^{LA} , updating U^{PCC} and U^{LA} , respectively;

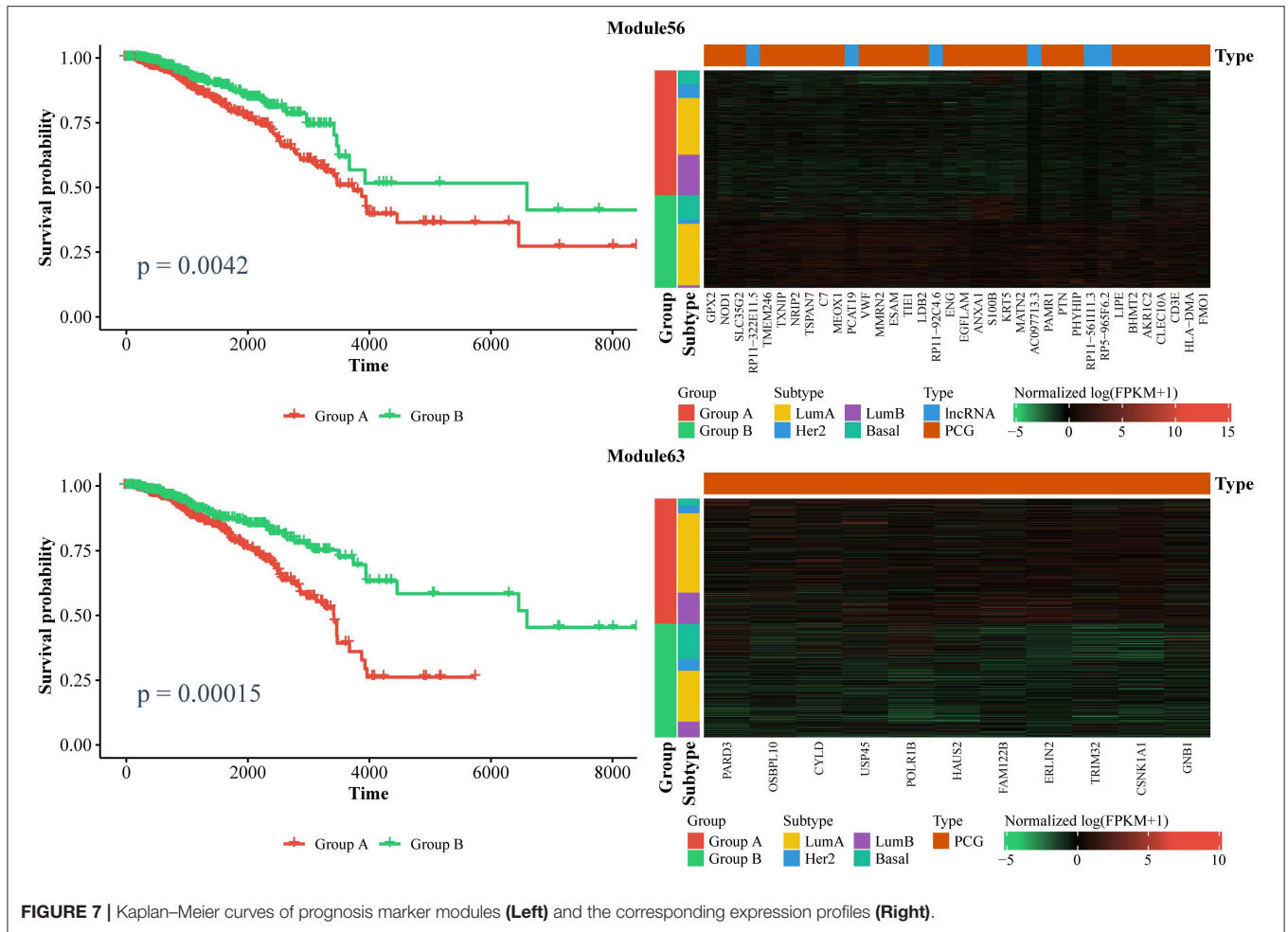


FIGURE 7 | Kaplan–Meier curves of prognosis marker modules (Left) and the corresponding expression profiles (Right).

$$U_{i,k}^{PCC} \leftarrow U_{i,k}^{PCC} \frac{(M^{PCC} V^{PCC})_{i,k} + \lambda_{PCC} \sum_{j=1}^N V_{j,k}^{PCC} V_{j,k}^*}{(U^{PCC} (V^{PCC})^T V^{PCC})_{i,k} + \lambda_{PCC} \sum_{l=1}^M U_{l,k}^{PCC} \sum_{j=1}^N (V_{j,k}^{PCC})^2}$$

$$U_{i,k}^{LA} \leftarrow U_{i,k}^{LA} \frac{(M^{LA} V^{LA})_{i,k} + \lambda_{LA} \sum_{j=1}^N V_{j,k}^{LA} V_{j,k}^*}{(U^{LA} (V^{LA})^T V^{LA})_{i,k} + \lambda_{LA} \sum_{l=1}^M U_{l,k}^{LA} \sum_{j=1}^N (V_{j,k}^{LA})^2}$$

2. fixing V^* , U^{PCC} , and U^{LA} , updating V^{PCC} and V^{LA} , respectively:

$$V_{j,k}^{PCC} \leftarrow V_{j,k}^{PCC} \frac{((M^{PCC})^T U^{PCC})_{j,k} + \lambda_{PCC} V_{j,k}^*}{(V^{PCC} (U^{PCC})^T U^{PCC})_{j,k} + \lambda_{PCC} V_{j,k}^{PCC}}$$

$$V_{j,k}^{LA} \leftarrow V_{j,k}^{LA} \frac{((M^{LA})^T U^{LA})_{j,k} + \lambda_{LA} V_{j,k}^*}{(V^{LA} (U^{LA})^T U^{LA})_{j,k} + \lambda_{LA} V_{j,k}^{LA}}$$

3. fixing V^{PCC} , V^{LA} , U^{PCC} , and U^{LA} , updating V^* :

$$V^* = \frac{\lambda_{PCC} V^{PCC} Q^{PCC} + \lambda_{LA} V^{LA} Q^{LA}}{\lambda_{PCC} + \lambda_{LA}},$$

where

$$Q^{PCC} = \text{Diag} \left(\sum_{i=1}^G U_{i,1}^{PCC}, \sum_{i=1}^G U_{i,2}^{PCC}, \dots, \sum_{i=1}^G U_{i,K}^{PCC} \right)$$

$$Q^{LA} = \text{Diag} \left(\sum_{i=1}^G U_{i,1}^{LA}, \sum_{i=1}^G U_{i,2}^{LA}, \dots, \sum_{i=1}^G U_{i,K}^{LA} \right)$$

More details can be found in Liu et al. (2013).

Disease-Associated Genes Filter

We retained the patients with both tumor and normal samples and used them to identify differentially expressed genes using the R package edgeR software (Robinson et al., 2010). We set genes with $FDR < 0.05, |\log FC| > 1$ as differentially expressed genes.

Apart from the differentially expressed genes, we investigated the BRCA-associated copy number variation (CNV) genes

and differentially methylated genes to further investigate the BRCA associated modules. This was based on the notion that the expression of these genes may be altered and reflect the ceRNA interactions. We calculated the level of copy number variation for each gene in each sample using GISTIC2.0 software (Mermel et al., 2011), and identified genes that did not equal two in >5% of samples as CNV genes.

For DNA methylation, we used Illumina 450K methylation data. We initially identified differentially methylated sites (DMS) using Limma (Ritchie et al., 2015). The sites, which showed statistical significance with a $FDR < 0.05$ and had mean methylation differences between disease and normal states >0.2 , were marked as DMS. Subsequently, we isolated the sites located in the regions between 2,000bp upstream and downstream of the start position of ceRNA candidate genes and mapped them to related genes. For genes related to more than one DMS and more than three-quarters of these DMSs exhibit the same direction of change in methylation, we set these genes as differential methylation genes (Kim et al., 2012).

DATA AVAILABILITY STATEMENT

The RNA-seq data, microRNA-seq data, CNV data and gene methylation data are from The Cancer Genome Atlas (TCGA). The source codes of LAcModule is available at <https://github.com/GaoLabXDU/LAcModule>.

AUTHOR CONTRIBUTIONS

XW and LG conceived the study. XW and YH designed the experiments. XW performed the experiments and analyses. All authors discussed the results and contributed to the final manuscript.

REFERENCES

- Ai, D., Li, X., Pan, H., Chen, J., Cram, J. A., and Xia, L. C. (2019). Explore mediated co-varying dynamics in microbial community using integrated local similarity and liquid association analysis. *BMC Genomics* 20:185. doi: 10.1186/s12864-019-5469-8
- Ala, U., Karreth, F. A., Bosia, C., Pagnani, A., Taulli, R., Léopold, V., et al. (2013). Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7154–7159. doi: 10.1073/pnas.1222509110
- Alibek, K., Kakpenova, A., Mussabekova, A., Sypabekova, M., and Karatayeva, N. (2013). Role of viruses in the development of breast cancer. *Infect. Agent. Cancer* 8:32. doi: 10.1186/1750-9378-8-32
- Amarante, M. K., and Watanabe, M. A. (2009). The possible involvement of virus in breast cancer. *J. Cancer Res. Clin. Oncol.* 135, 329–337. doi: 10.1007/s00432-008-0511-2
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Bates, J. P., Derakhshandeh, R., Jones, L., and Webb, T. J. (2018). Mechanisms of immune evasion in breast cancer. *BMC Cancer* 18:556. doi: 10.1186/s12885-018-4441-3
- Brooks, M. D., and Wicha, M. S. (2015). Tumor twitter: cellular communication in the breast cancer stem cell niche. *Cancer Discov.* 5, 469–471. doi: 10.1158/2159-8290.CD-15-0327

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61532014, 61672407, and 61432010), and the National Key R&D Program of China (Grant No. 2018YFC0910400).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00235/full#supplementary-material>

Supplementary Table 1 | Most frequently enriched KEGG pathways.

Supplementary Table 2 | Summary of BRCA-dysregulated genes.

Supplementary Table 3 | Top 15 most frequently shared microRNAs in more than three modules.

Supplementary File 1 | Heatmap of PCC and LA values of modules enriched for GO:0007155 in disease and normal states (top right: disease state, bottom left: normal state).

Supplementary File 2 | Heatmap of PCC and LA values of modules enriched for KEGG:04514 in disease and normal states (top right: disease state, bottom left: normal state).

Supplementary File 3 | Heatmap of PCC and LA values of modules enriched for KEGG:04510 in disease and normal states (top right: disease state, bottom left: normal state).

Supplementary File 4 | Profiles of differentially expressed genes.

Supplementary File 5 | BRCA-associated modules (Supplementary of Figure 5A).

Supplementary File 6 | Kaplan–Meier curves for individual genes in Figure 6C.

Supplementary File 7 | Kaplan–Meier curves of prognosis marker modules (left) and the corresponding expression profiles (right). (Supplementary of Figure 7).

Supplementary File 8 | Comparison of the results of different values for the parameter λ_{LA} .

Supplementary File 9 | Patients groups by the expression of let-7 family and the gene set in Figure 6.

Supplementary File 10 | The evaluation of Van der Waerden's method.

- Cavallaro, U., and Christofori, G. (2001). Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough. *Biochim. Biophys. Acta* 1552, 39–45. doi: 10.1016/s0304-419x(01)00038-5
- Chiu, S. C., Chung, H. Y., Cho, D. Y., Chan, T. M., Liu, M. C., Huang, H. M., et al. (2014). Therapeutic potential of microRNA let-7: tumor suppression or impeding normal stemness. *Cell Transplant.* 23, 459–469. doi: 10.3727/096368914X678418
- Chiu, Y. C., Hsiao, T. H., Chen, Y., and Chuang, E. Y. (2015). Parameter optimization for constructing competing endogenous RNA regulatory network in glioblastoma multiforme and other cancers. *BMC Genomics* 16 (Suppl. 4):S1. doi: 10.1186/1471-2164-16-S4-S1
- Chou, C. H., Chang, N. W., Shrestha, S., Hsu, S. D., Lin, Y. L., Lee, W. H., et al. (2016). miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* 44, D239–247. doi: 10.1093/nar/gkv1258
- Cunningham, H. D., Shannon, L. A., Calloway, P. A., Fassold, B. C., Dunwiddie, I., Vielhauer, G., et al. (2010). Expression of the C-C chemokine receptor 7 mediates metastasis of breast cancer to the lymph nodes in mice. *Transl. Oncol.* 3, 354–361. doi: 10.1593/tlo.10178
- Do, D., and Bozdog, S. (2018). Cancerin: a computational pipeline to infer cancer-associated ceRNA interaction networks. *PLoS Comput. Biol.* 14:e1006318. doi: 10.1371/journal.pcbi.1006318
- Dweep, H., and Gretz, N. (2015). miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods* 12:697. doi: 10.1038/nmeth.3485

- Figliuzzi, M., Marinari, E., and De Martino, A. (2013). MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. *Biophys. J.* 104, 1203–1213. doi: 10.1016/j.bpj.2013.01.012
- Fry, S. A., Robertson, C. E., Swann, R., and Dwek, M. V. (2016). Cadherin-5: a biomarker for metastatic breast cancer with optimum efficacy in oestrogen receptor-positive breast cancers with vascular invasion. *Br. J. Cancer* 114, 1019–1026. doi: 10.1038/bjc.2016.66
- Gendelman, R., Xing, H., Mirzoeva, O. K., Sarde, P., Curtis, C., Feiler, H. S., et al. (2017). Bayesian network inference modeling identifies TRIB1 as a novel regulator of cell-cycle progression and survival in cancer cells. *Cancer Res.* 77, 1575–1585. doi: 10.1158/0008-5472.CAN-16-0512
- Gunderson, T., and Ho, Y. Y. (2014). An efficient algorithm to explore liquid association on a genome-wide scale. *BMC Bioinform.* 15:371. doi: 10.1186/s12859-014-0371-5
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013
- Hu, J., Li, X., Guo, X., Guo, Q., Xiang, C., Zhang, Z., et al. (2017). The CCR2 3'UTR functions as a competing endogenous RNA to inhibit breast cancer metastasis. *J. Cell Sci.* 130, 3399–3413. doi: 10.1242/jcs.202127
- Hubert, L., and Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *Brit. J. Math. Stat. Psy.* 29, 190–241. doi: 10.1111/j.2044-8317.1976.tb00714.x
- Jiang, Y., Pan, Y., Rhea, P. R., Tan, L., Gagea, M., Cohen, L., et al. (2016). A sucrose-enriched diet promotes tumorigenesis in mammary gland in part through the 12-lipoxygenase pathway. *Cancer Res.* 76, 24–29. doi: 10.1158/0008-5472.CAN-14-3432
- Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W. O., Corcoran, M., Grandér, D., et al. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* 20, 440–446. doi: 10.1038/nsmb.2516
- Kakolyris, S., Fox, S. B., Koukourakis, M., Giatromanolaki, A., Brown, N., Leek, R. D., et al. (2000). Relationship of vascular maturation in breast cancer blood vessels to vascular density and metastasis, assessed by expression of a novel basement membrane component, LH39. *Br. J. Cancer* 82, 844–851. doi: 10.1054/bjoc.1999.1010
- Karreth, F. A., and Pandolfi, P. P. (2013). ceRNA Cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* 3, 1113–1121. doi: 10.1158/2159-8290.CD-13-0202
- Kim, D. H., and Wirtz, D. (2013). Focal adhesion size uniquely predicts cell migration. *FASEB J.* 27, 1351–1361. doi: 10.1096/fj.12-220160
- Kim, J. W., Kim, S. T., Turner, A. R., Young, T., Smith, S., Liu, W., et al. (2012). Identification of new differentially methylated genes that have potential functional consequences in prostate cancer. *PLoS ONE* 7:e48455. doi: 10.1371/journal.pone.0048455
- Lawson, J. S., Salmons, B., and Glenn, W. K. (2018). Oncogenic viruses and breast cancer: mouse mammary tumor virus (MMTV), bovine leukemia virus (BLV), human papilloma virus (HPV), and epstein-barr virus (EBV). *Front. Oncol.* 8:1. doi: 10.3389/fonc.2018.00001
- Le, T. D., Zhang, J., Liu, L., and Li, J. (2017). Computational methods for identifying miRNA sponge interactions. *Brief. Bioinform.* 18, 577–590. doi: 10.1093/bib/bbw042
- Li, K. C. (2002). Genome-wide coexpression dynamics: theory and application. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16875–16880. doi: 10.1073/pnas.252466999
- Li, K. C., Palotie, A., Yuan, S., Bronnikov, D., Chen, D., Wei, X., et al. (2007). Finding disease candidate genes by liquid association. *Genome Biol.* 8:R205. doi: 10.1186/gb-2007-8-10-r205
- List, M., Dehghani Amirabad, A., Kostka, D., and Schulz, M. H. (2019). Large-scale inference of competing endogenous RNA networks with sparse partial correlation. *Bioinformatics* 35, i596–i604. doi: 10.1093/bioinformatics/btz314
- Liu, J., Wang, C., Gao, J., and Han, J. (2013). “Multi-view clustering via joint nonnegative matrix factorization,” in *Proceedings of the 2013 SIAM International Conference on Data Mining: SIAM* (Austin, TX), 252–260. doi: 10.1137/1.9781611972832.28
- Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: a literature-based database for human oncogenes. *J. Genet. Genomics* 44, 119–121. doi: 10.1016/j.jgg.2016.12.004
- Maragkakis, M., Vergoulis, T., Alexiou, P., Reczko, M., Plomaritou, K., Gousis, M., et al. (2011). DIANA-microT web server upgrade supports fly and worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res.* 39, W145–W148. doi: 10.1093/nar/gkr294
- Martirosyan, A., Del Giudice, M., Bena, C. E., Pagnani, A., Bosia, C., and De Martino, A. (2019). Kinetic modelling of competition and depletion of shared miRNAs by competing endogenous RNAs. *Methods Mol. Biol.* 1912, 367–409. doi: 10.1007/978-1-4939-8982-9_15
- McClain, J. O., and Rao, V. R. (1975). Clustis: a program to test for the quality of clustering of a set of objects. *J. Mark. Res.* 12, 456–460.
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12:R41. doi: 10.1186/gb-2011-12-4-r41
- Milligan, G. W. (1981). A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika* 46, 187–199. doi: 10.1007/BF02293899
- Mullokanov, G., Baccarini, A., Ruza, A., Jayaprakash, A. D., Tung, N., Israelow, B., et al. (2012). High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* 9, 840–846. doi: 10.1038/nmeth.2078
- Okegawa, T., Pong, R. C., Li, Y., and Hsieh, J. T. (2004). The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochim. Pol.* 51, 445–457. doi: 10.18388/abp.2004_3583
- Otto, T., and Sicsinski, P. (2017). Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* 17, 93–115. doi: 10.1038/nrc.2016.138
- Paci, P., Colombo, T., and Farina, L. (2014). Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst. Biol.* 8:83. doi: 10.1186/1752-0509-8-83
- Pian, C., Zhang, G., Tu, T., Ma, X., and Li, F. (2018). LncCeRBase: a database of experimentally validated human competing endogenous long non-coding RNAs. *Database* 2018:bay061. doi: 10.1093/database/bay061
- Poliseno, L., Haimovic, A., Christos, P. J., Vega y Saenz de Miera, E. C., Shapiro, R., Pavlick, A., et al. (2011). Deletion of PTENP1 pseudogene in human melanoma. *J. Invest. Dermatol.* 131, 2497–2500. doi: 10.1038/jid.2011.232
- Poliseno, L., and Pandolfi, P. P. (2015). PTEN ceRNA networks in human cancer. *Methods* 77–78, 41–50. doi: 10.1016/j.yemeth.2015.01.013
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.1038/nature09144
- Qi, X., Zhang, D. H., Wu, N., Xiao, J. H., Wang, X., and Ma, W. (2015). ceRNA in cancer: possible functions and clinical implications. *J. Med. Genet.* 52, 710–718. doi: 10.1136/jmedgenet-2015-103334
- Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 35, W193–W200. doi: 10.1093/nar/gkm226
- Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., et al. (2003). Cell migration: integrating signals from front to back. *Science* 302, 1704–1709. doi: 10.1126/science.1092053
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014
- Schmitt, A. M., and Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell* 29, 452–463. doi: 10.1016/j.ccell.2016.03.010
- Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* 9, 1135–1151. doi: 10.1214/aos/1176345632
- Strehl, A., and Ghosh, J. (2003). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* 3, 583–617. doi: 10.1162/153244303321897735

- Sumazin, P., Yang, X., Chiu, H. S., Chung, W. J., Iyer, A., Llobet-Navas, D., et al. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147, 370–381. doi: 10.1016/j.cell.2011.09.041
- Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352. doi: 10.1038/nature12986
- Thomson, D. W., and Dinger, M. E. (2016). Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* 17, 272–283. doi: 10.1038/nrg.2016.20
- Tomita, Y., Dorward, H., Yool, A. J., Smith, E., Townsend, A. R., Price, T. J., et al. (2017). Role of aquaporin 1 signalling in cancer development and progression. *Int. J. Mol. Sci.* 18:99. doi: 10.3390/ijms18020299
- Wang, B., Liu, M., Song, Y., Li, C., Zhang, S., and Ma, L. (2019). KLF2 inhibits the migration and invasion of prostate cancer cells by downregulating MMP2. *Am. J. Mens. Health* 13:1557988318816907. doi: 10.1177/1557988318816907
- Wang, H. G., Cao, B., Zhang, L. X., Song, N., Li, H., Zhao, W. Z., et al. (2017). KLF2 inhibits cell growth via regulating HIF-1 α /Notch-1 signal pathway in human colorectal cancer HCT116 cells. *Oncol. Rep.* 38, 584–590. doi: 10.3892/or.2017.5708
- Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., et al. (2010). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* 38, 5366–5383. doi: 10.1093/nar/gkq285
- Wang, P., Gao, L., Hu, Y., and Li, F. (2018). Feature related multi-view nonnegative matrix factorization for identifying conserved functional modules in multiple biological networks. *BMC Bioinform.* 19:394. doi: 10.1186/s12859-018-2434-5
- Wang, P., Ning, S., Zhang, Y., Li, R., Ye, J., Zhao, Z., et al. (2015a). Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res.* 43, 3478–3489. doi: 10.1093/nar/gkv233
- Wang, P., Zhi, H., Zhang, Y., Liu, Y., Zhang, J., Gao, Y., et al. (2015b). miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. *Database.* 2015:bav098. doi: 10.1093/database/bav098
- Wang, X., and El Naqa, I. M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics* 24, 325–332. doi: 10.1093/bioinformatics/btm595
- Wei, L., Yuan, Y., Hu, T., Li, S., Cheng, T., Lei, J., et al. (2019). Regulation by competition: a hidden layer of gene regulatory network. *J. Quant. Biol.* 7, 110–121. doi: 10.1007/s40484-018-0162-5
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764
- Yang, J., Li, T., Gao, C., Lv, X., Liu, K., Song, H., et al. (2014). FOXO1 3'UTR functions as a ceRNA in repressing the metastases of breast cancer cells via regulating miRNA activity. *FEBS Lett.* 588, 3218–3224. doi: 10.1016/j.febslet.2014.07.003
- Yang, J., Qin, Y., Zhang, T., Wang, F., Peng, L., Zhu, L., et al. (2018). Identification of human age-associated gene co-expressions in functional modules using liquid association. *Oncotarget* 9, 1063–1074. doi: 10.18632/oncotarget.23148
- Yates, L. A., Norbury, C. J., and Gilbert, R. J. (2013). The long and short of microRNA. *Cell* 153, 516–519. doi: 10.1016/j.cell.2013.04.003
- Yu, G., Yao, W., Gumireddy, K., Li, A., Wang, J., Xiao, W., et al. (2014). Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma progression. *Mol. Cancer Ther.* 13, 3086–3097. doi: 10.1158/1535-7163.MCT-14-0245
- Yu, T. (2018). A new dynamic correlation algorithm reveals novel functional aspects in single cell and bulk RNA-seq data. *PLoS Comput. Biol.* 14:e1006391. doi: 10.1371/journal.pcbi.1006391
- Yuan, Y., Liu, B., Xie, P., Zhang, M. Q., Li, Y., Xie, Z., et al. (2015). Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3158–3163. doi: 10.1073/pnas.1413896112
- Zhang, J., Liu, L., Li, J., and Le, T. D. (2018). lncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer. *Bioinformatics* 34, 4232–4240. doi: 10.1093/bioinformatics/bty525
- Zhang, W., Levi, L., Banerjee, P., Jain, M., and Noy, N. (2015). Kruppel-like factor 2 suppresses mammary carcinoma growth by regulating retinoic acid signaling. *Oncotarget* 6, 35830–35842. doi: 10.18632/oncotarget.5767
- Zheng, L., Zhang, Z., Zhang, S., Guo, Q., Zhang, F., Gao, L., et al. (2018). RNA binding protein RNPC1 inhibits breast cancer cell metastasis via activating STARD13-correlated ceRNA network. *Mol. Pharm.* 15, 2123–2132. doi: 10.1021/acs.molpharmaceut.7b01123
- Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., et al. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 7:11215. doi: 10.1038/ncomms11215
- Zhong, Y., Du, Y., Yang, X., Mo, Y., Fan, C., Xiong, F., et al. (2018). Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer.* 17:79. doi: 10.1186/s12943-018-0827-8
- Zhou, X., Liu, J., and Wang, W. (2014). Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. *IET Syst. Biol.* 8, 96–103. doi: 10.1049/iet-syb.2013.0025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wen, Gao and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.