Identification of Four Serum miRNAs as Potential Markers to Screen for Thirteen Cancer Types

Joe W. Chen
California University of Science and Medicine
Joseph Dhahbi (dhahbij@cusm.org)
California University of Science and Medicine

Research Article

Keywords: Microarray, microRNA, miRNA, biomarker, cancer screening, serum

Posted Date: August 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-783019/v1

Abstract

Introduction: Cancer has consistently remained one of the top causes of death in the United States every year, with many cancer deaths preventable if detected early. Circulating serum miRNAs pose as a promising, minimally invasive supplement, or even alternative, to many current screening procedures. Many studies have shown that different serum miRNAs can discriminate healthy individuals from those with certain types of cancer. Although many of those miRNAs are often reported to be significant in one cancer type, they are also altered in other types of cancers. Currently, very few studies have investigated serum miRNA biomarkers for multiple different cancer types simultaneously for general cancer screening.

Method: To identify serum miRNAs that would be useful in simultaneously screening multiple types of cancers, microarray cancer datasets were curated, yielding 13 different types of cancer with a total of 3352 cancer samples and 2809 non-cancer samples. The samples were then divided into discovery and validation sets. A hundred random forest models were built using the discovery set to select candidate miRNAs. The selected miRNAs were then used in the validation set to see how well they can differentiate cancer from normal samples in an independent dataset. Further analysis of the interactions between these miRNAs and their target mRNAs were investigated as well.

Result: The random forest models achieved an average of 97% accuracy in the discovery set with 95% bootstrap confidence interval from 0.9544 to 0.9778. The selected miRNAs were hsa-miR-663a, hsa-miR-6802-5p, hsa-miR-6784-5p, hsa-miR-3184-5p, and hsa-miR-8073. Each individual miRNA exhibited high area under the curve (AUC) value using receiver operating characteristic analysis. Moreover, the combination of four out of the five miRNAs achieved the highest AUC value of 0.9815 with perfect sensitivity, indicating that these miRNAs have a high potential for cancer screening. miRNA-mRNA interaction and protein-protein interaction analysis provided insights into how these miRNAs may play a role in cancer in general.

Introduction

Cancer has consistently been one of the most common causes of death in the United States, precisely the second leading cause in 2020 [1]. Therefore, effective cancer screening and early detection of cancer are crucial for improving healthcare outcomes [2, 3]. However, many of the current standards for cancer screening lack sufficient sensitivity and specificity, and many of the screening modalities are invasive [3]. In addition, many cancers such as ovarian cancer and pancreatic cancer are known to be deadly because of late-stage discoveries [4, 5]. With the stable nature of miRNAs, circulating serum miRNAs can serve as a minimally invasive alternative or supplement the current standard for cancer screening [6, 7].

Many miRNAs have already been reported to be promising biomarkers for certain types of cancer. For instance, plasma miR-145, miR-20a, miR-21, and miR-223 have been shown to be biomarkers for screening of early-stage non-small cell lung cancer [8, 9]. Similarly, serum miR-21 and other miRNAs are also found to be differentially regulated in glioma compared to healthy controls [10]. Other different

panels of miRNAs can be used in early-stage breast, colorectal, and other cancer diagnoses as well [11–18]. However, some of these reported miRNAs may be non-specific; many miRNAs while being reported to be important in one cancer are also altered in other types of cancers. Rarely, has any study investigated serum miRNAs for multiple different types of cancers for general cancer screening[19–21].

In this study, we curated large microarray datasets consisting of different types of cancers along with non-cancer samples. The cancers include breast, lung, colorectal, prostate, and gastric cancers, which are the top five most prevalent cancers in the world in 2020 [22]. The curated dataset also comprises ovarian and pancreatic cancers, which are well-known to present in late stages [4, 5]. In addition, the dataset includes biliary tract cancers, bladders cancer, liver cancers, esophageal cancers, gliomas, and sarcomas [14, 16–18]. Candidate miRNAs for general cancer screening for these 13 types of cancers were selected via random forest, a widely used and reliable machine learning algorithm for biomarker discovery [23]. The selected miRNAs were then validated in an independent validation set, and a multinomial logistic regression model was built to distinguish cancer from non-cancer samples. We investigated the miRNA-mRNA interactions and enriched biological pathways to elucidate the roles these miRNAs may play in cancers.

Results

Study Design

The curated dataset from the four GEO datasets [14, 16–18] yielded 13 different types of cancers and many non-cancer samples (Table S1). There was a total of 3352 cancer samples and 2809 non-cancer samples. The clinical information of all the samples is detailed in Table 1. The curated dataset was split into discovery set with 2253 cancer samples and 2247 non-cancer samples, and the validation set with 1102 cancer samples and 562 non-cancer samples. We used the discovery set to select stable miRNAs via 100 random forest models and the validation set to verify the selected miRNA as a potential diagnostic marker for cancer detection. The selected miRNAs were then used to perform miRNA-mRNA network analysis, protein-protein interaction clustering analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis [24] (Fig. 1).

Table 1

Clinical Summary for Cancer Samples. The pathological stage and grade of many samples were unknown. Many of patients' age from which the sample was obtained were unknown. Only 90 samples from sarcomas, colorectal, esophageal, pancreatic, and gastric cancers were used to calculate mean and standard deviation of age. Only 40 samples from lung and breast cancer were used to calculate mean and standard deviation of age.

Cancer type	Mean Age (years)	Standard Deviation (Age in years)	Number of Female:Male	Total Number of Samples	Pathological Stage (if known)
Prostate	67.6	7.5	0:809	809	
Ovarian	56.9	11.5	320:0	320	
Bladder	67.8	10.7	109:283	392	313 with high pathological grade and 77 low pathological grade
Hepatocellular Carcinoma	67.6	9.2	77:268	345	270 child-pugh A, 34 child-pugh B
Breast	55.9	11	155:0	155	
Colorectal	65.7	11	130:75	205	
Sarcoma	53.0	17.6	133:72	205	
Pancreatic	63.5	10.0	130:75	205	
Gastric	66.7	10.0	124:81	205	
Lung	63.8	8.4	126:27	153	
Esophageal	67.3	8.2	119:86	178	
Glioma	52.6	18.6	21:69	90	
Biliary	67.7	9.4	12:78	90	

miRNAs Selected as Candidate Biomarkers and Heatmap Analysis

Five miRNAs were considered "stable", as they satisfied the criteria of being in the top 10 miRNAs 90% of the time across 100 random forest models [23, 25]. The miRNAs selected were: hsa-miR-663a, hsa-miR-6802-5p, hsa-miR-6784-5p, hsa-miR-3184-5p, and hsa-miR-8073 (Table 2). The random forest models achieved an average of 97% accuracy with 95% bootstrap confidence interval of 0.9544 to 0.9778 [26]. The hierarchical heatmap clustering was performed using these 5 miRNAs across the samples, and the heatmap shows a clear separation between cancer and non-cancer samples [27] (Fig. 2).

Table 2

Frequency of miRNAs. The table shows the frequency of the miRNAs in the top miRNAs displayed in the 100 random forest models. The top 5 miRNAs in this table were chosen to be biomarker candidates for cancer screening.

miRNA	Frequency
Hsa-miR-3184-5p	100
Hsa-miR-663a	100
Hsa-miR-6784-5p	100
Hsa-miR-6802-5p	96
Hsa-miR-8073	90
Hsa-miR-4783-3p	87
Hsa-miR-1307-3p	86
Hsa-miR-4730	79
Hsa-miR-320a-3p	63
Hsa-miR-5100	45
Hsa-miR-1343-3p	43
Hsa-miR-1469	38
Hsa-miR-1233-5p	30
Hsa-miR-1290	14
Hsa-miR-4675	11
Hsa-miR-1238-5p	8
Hsa-miR-320b	7
Hsa-miR-4532	2
Hsa-miR-4687-5p	1

miRNA AUC Validation and Generation of Diagnostic Models for Cancer Screening

The receiver operating characteristic curves and the area under the curve value were used to evaluate the diagnostic potential of each individual miRNA and their combinations both in the discovery set and the

validation set [28] (Fig. 3). Each of the miRNAs showed significant AUC values as displayed in Fig. 3. The best combination model used only four of the miRNAs: hsa-miR-663a, hsa-miR-6802-5p, hsa-miR-3184-5p, and hsa-miR-8073. The combined model was built using the discovery set, yielding (0.0005032411) x hsa-miR-663a + (0.0006917428) x hsa-miR-6802-5p + (0.0072807475) x hsa-miR-8073 + (-0.0194274974) x hsa-miR-3184-5p + (-1.1271024323) with AUC value of 0.9742 in the discovery set. The same model was then used to predict the cancer samples in the validation set, resulting in an accuracy of 0.9886, sensitivity of 1, specificity of 0.9673, and an AUC value of 0.9815.

miRNA-mRNA Network and Functional Enrichment Analysis

The network (Fig. 4) generated a total of 535 mRNAs, with many of them directly associated with cancer [29]. KEGG analysis of the generated mRNAs was performed. It yielded many significant pathways associated with cancer as well [24] (Table 3), with the most significant one being cell cycle and the second being chronic myeloid leukemia, along with many other cancers, including but not limited to glioma, prostate cancer, bladder cancer, and others. KEGG analysis using the mRNAs and circular RNAs yielded similar results (Table S2).

Table 3
KEGG Analysis Using mRNAs Associated with the 5 Selected miRNAs

Pathways	Hits	P value	Adj. P-value
Cell cycle	12	0.00022	0.01135
Chronic myeloid leukemia	9	0.000227	0.01135
Lysine degradation	7	0.000357	0.0119
Glioma	8	0.000517	0.012925
Neurotrophin signaling pathway	10	0.00282	0.05
p53 signaling pathway	7	0.00331	0.05
Prostate cancer	8	0.0035	0.05
Bladder cancer	4	0.00922	0.11525
HTLV-I infection	12	0.0123	0.1227273
Leukocyte transendothelial migration	8	0.0127	0.1227273
Melanoma	6	0.0135	0.1227273
Fructose and mannose metabolism	4	0.0196	0.1557143
Osteoclast differentiation	8	0.0217	0.1557143
Alcoholism	10	0.0218	0.1557143
Circadian rhythm - mammal	3	0.0247	0.1646667
Endometrial cancer	4	0.0378	0.2347059
ErbB signaling pathway	6	0.0399	0.2347059
MAPK signaling pathway	13	0.0429	0.2383333
Phototransduction	3	0.0463	0.24
Epstein-Barr virus infection	6	0.048	0.24

Protein-Protein Clustering Analysis

The Protein-Protein-Interaction (PPI) network from the top three resulting clusters yielded 46 nodes, 202 edges, and an average node degree of 8.78 [30]. The PPI enrichment p-value is < 1.0e-16, and KEGG analysis of these proteins yielded many similar pathways with more significant p-values than the previous KEGG analyses (Table S3). The clusters, along with their interaction with the five chosen miRNAs, are displayed in Fig. 5.

Discussion

Many studies have reported specific upregulation or downregulation of serum miRNAs in certain types of cancers [9, 11–18, 31–35]. Yet, few studies have investigated the potential of serum miRNAs as a general cancer screening markers across multiple cancer subtypes [6, 19]. To our knowledge, this is the largest study with the aim to assess the potential of miRNAs as markers for general cancer screening.

There are a total of 19 miRNAs identified by random forest models to be important in cancer diagnosis (Table 2); however, only five stable miRNAs were selected: hsa-miR-663a, hsa-miR-6802-5p, hsa-miR-6784-5p, hsa-miR-3184-5p, and hsa-miR-8073. Consistent with our result, hsa-miR-663a has been reported to regulate cancer signaling and tumor progression; it specifically has been shown to be a sensitive circulating miRNA marker for early detection of hepatocellular carcinoma [14, 36]. Similarly, hsa-miR-6784-5p has been reported to be a sensitive serum biomarker for early ovarian cancer diagnosis and a key regulator for breast cancer [37]. Hsa-miR-3184-5p is also a key regulator in breast cancer and a good biomarker for early detection of bladder cancer [18, 38]. Interestingly, hsa-miR-8073 is a natural tumor suppressor and identified to be a promising serum biomarker for ovarian and pancreatic cancers [37, 39]. These consistent and overlapping results suggest that these miRNAs may serve as cancer suitable for screening purposes.

Indeed, each of the five miRNAs achieved remarkable results with AUC values well over 0.9 in both discovery and validation sets. This indicates that the five miRNAs can function as a stand-alone diagnostic marker for at least the 13 types of cancers included in this study, in which some are known for late-stage presentations [4, 5]. Furthermore, the combined model of using four miRNAs: (0.0005032411) x hsa-miR-663a + (0.0006917428) x hsa-miR-6802-5p + (0.0072807475) x hsa-miR-8073 + (-0.0194274974) x hsa-miR-3184-5p + (-1.1271024323) achieved the highest AUC value of 0.9815 with perfect sensitivity, which is highly desirable for screening [2].

miRNA-mRNA network (Fig. 4) based on these five miRNAs also further provide evidence that these miRNAs are generally associated with cancers. These miRNAs target many cancer-associated genes including, TP53, ABL1, STAT5B, and E2F3 [40–43] (Fig. 4). KEGG analyses also show many enriched cancer-related pathways such as cell cycle, chronic myeloid leukemia, glioma, neurotrophin-signaling, and more [44–46] (Table 3, Table S2, Table S3). The top three clusters of mRNAs in the PPI network analysis also pointed toward a very similar result (Fig. 5). Many of the same cancer pathways showed up in KEGG with more significant p-values, indicating that these clustered mRNAs are the main actors in enriching these cancer pathways.

There are few limitations to this study. First, the five proposed serum miRNAs have yet to be independently verified. Second, many of the cancer subtypes and clinical information for cancer and non-cancer samples are unknown, contributing to possible biases. Finally, despite the effort to balance cancer samples while building the cancer diagnosis model, there are still some imbalances in the number of different cancer types while constructing the model, which may over-represent one cancer over the other.

Nevertheless, the results show perfect sensitivity and high AUC value for the proposed 4-miRNA panel. Each individual miRNA achieved significant diagnostic potentials, suggesting that these miRNAs can be used as minimally invasive biomarkers for general cancer screening. Moreover, network and KEGG analyses provided insights into how these miRNAs may play a role in cancer regulation; however, functional studies of these miRNAs and their associated mRNAs are warranted.

Method

Microarray Data Processing

GSE113740, GSE112264, GSE106817, GSE113486 datasets were obtained from GEO [14, 16–18]. These datasets were part of the Japan Initiative to sequence cancer transcriptome via microarray. All the datasets used in the study are publicly available as open source. In each dataset, the presence of miRNA was determined if the signal was greater than the mean + 2 X standard deviation of the internal negative control. The background signals were subtracted, and samples were quantile normalized. The series matrix files were downloaded. Each dataset contains different cancer types with some overlapping samples. Therefore, we manually curated the four datasets to ensure that there are no duplicated samples while maximizing the number of samples. The curated data contain 13 different types of cancers. The distribution of cancers and the clinical information of the dataset are provided in Table 1. We randomly separated the curated data into discovery set and validation set before analysis by computer-generated random numbers. To minimize bias over certain cancer types with more samples, we randomly chose 40% of prostate cancers and 70% of ovarian, liver cancers, and bladder cancers to be in the discovery set. We randomly chose 80% of the samples in the rest of the cancer types to include in the discovery set. The remaining samples made up the samples in the validation set. Then we used the discovery set to select miRNAs that can successfully screen out cancer samples from normal samples, and the validation set to validate the result. The workflow of this study is provided in Fig. 1.

Stable miRNA Selection as Candidate Biomarkers

We grouped the different cancer types as cancer samples and compared their expression values to those of the non-cancer samples within the discovery set. Welch's t-test was performed for each miRNA, and FDR was calculated. Top 500 miRNAs with the least FDR were chosen to undergo further selection via random forest. A total of 100 random forest models were built using R and the ranger package [25], with each model randomly selecting 80% of the discovery set for training, and the remaining 20% for testing. Mean decrease in Gini indices, or the importance value for the 500 miRNAs, were obtained for each model. The miRNAs were deemed "stable" if they show up as one of the top 10 miRNAs in their importance values for over 90 models. Bootstrapping in R using 10,000 replicates was used to calculate the confidence interval of the accuracy achieved by the random forest model [26]. We plotted the hierarchical heatmap of expression values for these miRNAs to show separation between cancer and non-cancers [27] (Fig. 2).

miRNA Cancer Classification and Validation

We used receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC) to evaluate each individual miRNA's potential in distinguishing cancer from normal samples [28]. The curves were generated for both discovery and validation sets (Fig. 3). A higher AUC value indicates a higher distinguishing potential for the miRNA. To improve the discriminating potential even further, multinomial logistic regression model was used to discriminate cancer from the non-cancer samples using combinations of the miRNAs [47].

MiRNA-mRNA Interaction and Functional Enrichment Analysis

To further study why and how these miRNAs are important in cancer diagnosis, we used miRnet [29] to analyze the relationship between the chosen miRNAs and their associated mRNAs. miRnet [29] is a web-based software that displays all the miRNA-mRNA interactions, providing insight into how these miRNAs might regulate different mRNAs associated with cancer [29]. The analysis was performed with setting organism into homo sapiens and unspecified tissue of origin. Furthermore, two KEGG analyses were performed – one using all the associated mRNAs (Table 2) and one with all the circular RNAs in addition to the mRNAs [24, 48] (Table S2).

Protein-protein Interaction Cluster Analysis

The target genes from the miRNA-mRNA interaction network were further clustered to uncover their potential contribution to the development of cancer [29]. The miRNA-mRNA network was uploaded and visualized in the Cytoscape software [49]. Then, the top clusters were chosen using the Molecular Complex Detection (MCODE) technique, with the inclusion criteria of degree cutoff of 2, node score cutoff of 2, k-core of 2, and the maximum depth of 100 [50]. The threshold MCODE score was set to greater or equal to 5 as criteria. The resulting clusters were plotted together as a network; also, the mRNAs were input into the STRING database v 11 to visualize their interactions [30]. KEGG analysis was performed using the proteins from the selected clusters as well [24] (Table S3).

Declarations

Authors' contributions

JC proposed the study and obtained, analyzed, and interpreted the results. JC wrote the manuscript and generated the figures and tables. JD supervised and revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

All data generated and analyzed during the current study are included in this published article (and its supplementary information files). The custom code used for data analysis can be accessed at https://github.com/chenjoe569/CancerScreenResearch. All generated datasets can be produced using the custom code. For convenience, all generated datasets will be available upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

References

1. Ahmad, F.B. and R.N. Anderson, *The Leading Causes of Death in the US for 2020.* JAMA, 2021. **325**(18): p. 1829-1830.
2. Iragorri, N. and E. Spackman, *Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis.* Public Health Rev, 2018. **39**: p. 17.
3. Duffy, M.J., *Use of Biomarkers in Screening for Cancer.* EJIFCC, 2010. **21**(1): p. 1-12.
4. Mizrahi, J.D., et al., *Pancreatic cancer.* Lancet, 2020. **395**(10242): p. 2008-2020.
5. Stewart, C., C. Ralyea, and S. Lockwood, *Ovarian Cancer: An Integrated Review.* Semin Oncol Nurs, 2019. **35**(2): p. 151-156.
6. Wang, H., et al., *Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage.* Clin Epigenetics, 2018. **10**: p. 59.
7. Galvao-Lima, L.J., et al., *miRNAs as biomarkers for early cancer detection and their application in the development of new diagnostic tools.* Biomed Eng Online, 2021. **20**(1): p. 21.
8. Zhang, H., et al., *Plasma miR-145, miR-20a, miR-21 and miR-223 as novel biomarkers for screening early-stage non-small cell lung cancer.* Oncol Lett, 2017. **13**(2): p. 669-676.
9. Geng, Q., et al., *Five microRNAs in plasma as novel biomarkers for screening of early-stage non-small cell lung cancer.* Respir Res, 2014. **15**: p. 149.
10. Santangelo, A., et al., *Circulating microRNAs as emerging non-invasive biomarkers for gliomas.* Ann Transl Med, 2017. **5**(13): p. 277.
11. Zhu, M., et al., *A panel of microRNA signature in serum for colorectal cancer diagnosis.* Oncotarget, 2017. **8**(10): p. 17081-17091.
12. Fu, L., et al., *Serum expression levels of microRNA-382-3p, -598-3p, -1246 and -184 in breast cancer patients.* Oncol Lett, 2016. **12**(1): p. 269-274.
13. Jiang, Z., et al., *Serum microRNA-218 is a potential biomarker for esophageal cancer.* Cancer Biomark, 2015. **15**(4): p. 381-9.
14. Yamamoto, Y., et al., *Highly Sensitive Circulating MicroRNA Panel for Accurate Detection of Hepatocellular Carcinoma in Patients With Liver Disease.* Hepatol Commun, 2020. **4**(2): p. 284-297.
15. So, J.B.Y., et al., *Development and validation of a serum microRNA biomarker panel for detecting gastric cancer in a high-risk population.* Gut, 2021. **70**(5): p. 829-837.
16. Urabe, F., et al., *Large-scale Circulating microRNA Profiling for the Liquid Biopsy of Prostate Cancer.* Clin Cancer Res, 2019. **25**(10): p. 3016-3025.
17. Yokoi, A., et al., *Integrated extracellular microRNA profiling for ovarian cancer screening.* Nat Commun, 2018. **9**(1): p. 4319.

18. Usuba, W., et al., *Circulating miRNA panels for specific and early detection in bladder cancer.* Cancer Sci, 2019. **110**(1): p. 408-419.
19. Lodes, M.J., et al., *Detection of cancer with serum miRNAs on an oligonucleotide microarray.* PLoS One, 2009. **4**(7): p. e6229.
20. Saliminejad, K., H.R. Khorram Khorshid, and S.H. Ghaffari, *Why have microRNA biomarkers not been translated from bench to clinic?* Future Oncol, 2019. **15**(8): p. 801-803.
21. Brase, J.C., et al., *Serum microRNAs as non-invasive biomarkers for cancer.* Mol Cancer, 2010. **9**: p. 306.
22. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. **71**(3): p. 209-249.
23. Acharjee, A., et al., *A random forest based biomarker discovery and power analysis framework for diagnostics research.* BMC Med Genomics, 2020. **13**(1): p. 178.
24. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
25. Wright MN, Z.A., *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R Journal of Statistical Software.* Journal of Statistical Software, 2017. **77**(1): p. 1-17.
26. Canty A, R.B., *boot: Bootstrap R (S-plus) Functions.* 2020.
27. Zhao, S.G., Y., Sheng, Q. et al., *Heatmap3: an improved heatmap package with more powerful and convenient features.* BMC Bioinformatics, 2014. **15**.
28. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves.* BMC Bioinformatics, 2011. **12**: p. 77.
29. Chang, L., et al., *miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology.* Nucleic Acids Res, 2020. **48**(W1): p. W244-W251.
30. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
31. Han, Y., et al., *Serum microRNAs as Biomarkers for the Noninvasive Early Diagnosis of Biliary Tract Cancer.* Int J Gen Med, 2021. **14**: p. 1185-1195.
32. Yang, X., et al., *Serum microRNA Signature Is Capable of Early Diagnosis for Non-Small Cell Lung Cancer.* Int J Biol Sci, 2019. **15**(8): p. 1712-1722.
33. Jin, Z., et al., *A novel serum miRNA-pair classifier for diagnosis of sarcoma.* PLoS One, 2020. **15**(7): p. e0236097.
34. Khan, I.A., et al., *Panel of serum miRNAs as potential non-invasive biomarkers for pancreatic ductal adenocarcinoma.* Sci Rep, 2021. **11**(1): p. 2824.
35. Yu, Y., et al., *Identification of Serum microRNA-25 as a novel biomarker for pancreatic cancer.* Medicine (Baltimore), 2020. **99**(52): p. e23863.

36. Zhang, C., et al., *miR-663a inhibits tumor growth and invasion by regulating TGF-beta1 in hepatocellular carcinoma.* BMC Cancer, 2018. **18**(1): p. 1179.
37. Yaghoobi, H., et al., *EBST: An Evolutionary Multi-Objective Optimization Based Tool for Discovering Potential Biomarkers in Ovarian Cancer.* IEEE/ACM Trans Comput Biol Bioinform, 2020. **PP**.
38. Rajarajan, D., et al., *Genome-wide analysis reveals miR-3184-5p and miR-181c-3p as a critical regulator for adipocytes-associated breast cancer.* J Cell Physiol, 2019. **234**(10): p. 17959-17974.
39. Shams, R., et al., *Identification of potential microRNA panels for pancreatic cancer diagnosis using microarray datasets and bioinformatics methods.* Sci Rep, 2020. **10**(1): p. 7559.
40. Petitjean, A., et al., *TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes.* Oncogene, 2007. **26**(15): p. 2157-65.
41. Greuber, E.K., et al., *Role of ABL family kinases in cancer: from leukaemia to solid tumours.* Nat Rev Cancer, 2013. **13**(8): p. 559-71.
42. Rani, A. and J.J. Murphy, *STAT5 in Cancer and Immunity.* J Interferon Cytokine Res, 2016. **36**(4): p. 226-37.
43. Feng, Z., et al., *E2F3 promotes cancer growth and is overexpressed through copy number variation in human melanoma.* Onco Targets Ther, 2018. **11**: p. 5303-5313.
44. Griffin, N., et al., *Targeting neurotrophin signaling in cancer: The renaissance.* Pharmacol Res, 2018. **135**: p. 12-17.
45. Chen, L., et al., *Pan-Cancer Analysis Reveals the Functional Importance of Protein Lysine Modification in Cancer Development.* Front Genet, 2018. **9**: p. 254.
46. Collins, K., T. Jacks, and N.P. Pavletich, *The cell cycle and cancer.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 2776-8.
47. Venables WN, R.B., *Modern Applied Statistics with S, Fourth edition*. 2002, New York: Springer.
48. Greene, J., et al., *Circular RNAs: Biogenesis, Function and Role in Human Diseases.* Front Mol Biosci, 2017. **4**: p. 38.
49. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
50. Saito, R., et al., *A travel guide to Cytoscape plugins.* Nat Methods, 2012. **9**(11): p. 1069-76.

Figures

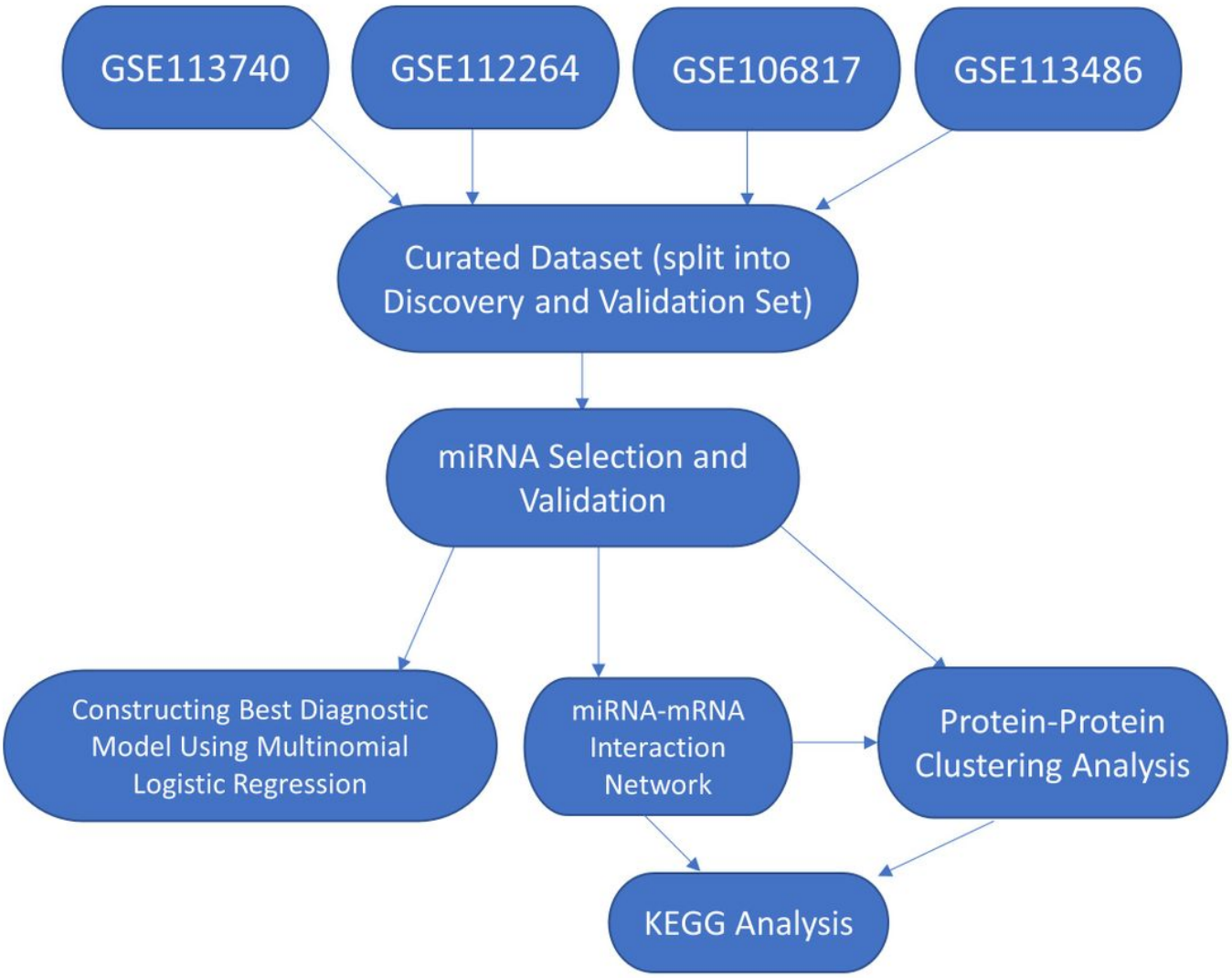

Figure 1

Analytical workflow. The microarray data were manually curated from four studies and combined for miRNA selection. One sentence about selection. The selected miRNAs were then used to classify and validate cancer subjects. miRNA-mRNA interaction network, protein-protein clustering analysis, and KEGG analysis were performed as well.

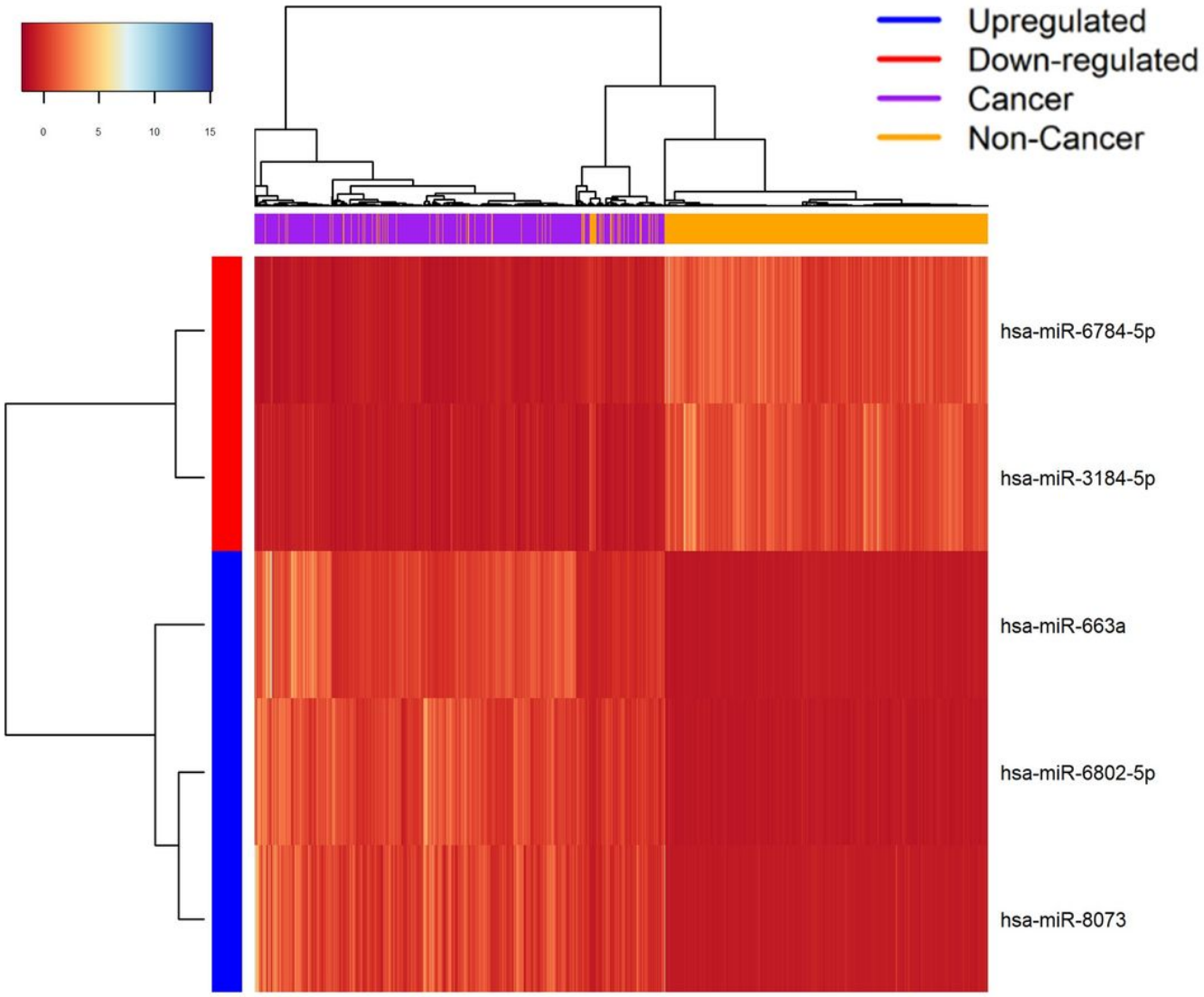

Figure 2

Heatmap of the expression value of the top 5 selected miRNAs. The X-axis represents the samples, and the Y-axis represents the miRNAs. Each of the boxes represents the normalized expression value of each miRNA in the corresponding sample.

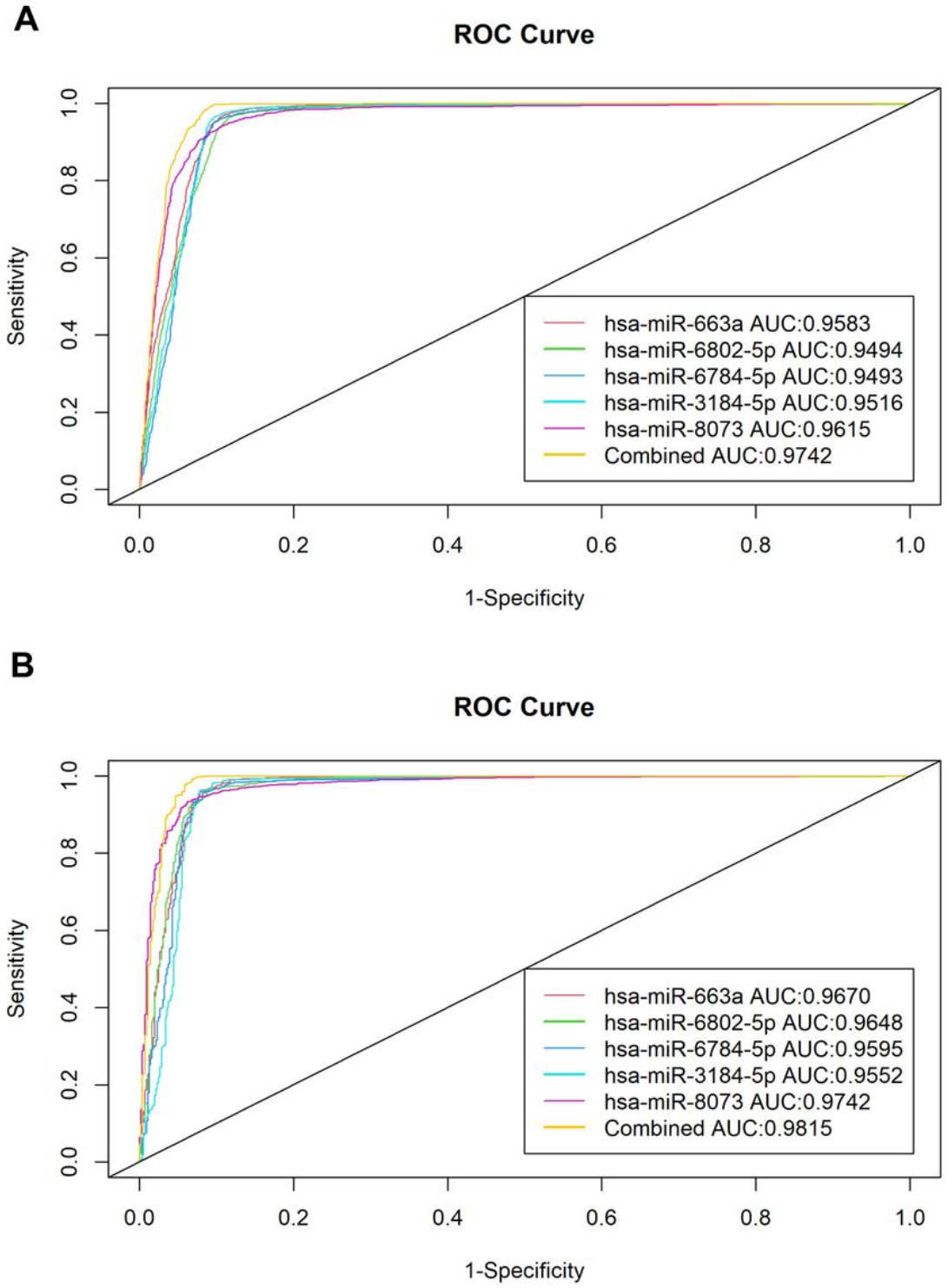

Figure 3

ROC and AUC Analysis of the top 5 selected miRNAs and their best combination ROC and AUC values. Panel A is the analysis for the discovery set. Panel B is the analysis for the Validation Set. Both panels achieved the highest ROC and AUC value using 4 miRNAs: has-miR-663a, has-miR-6802, has-miR-3184-5p, and hsa-miR-8073.

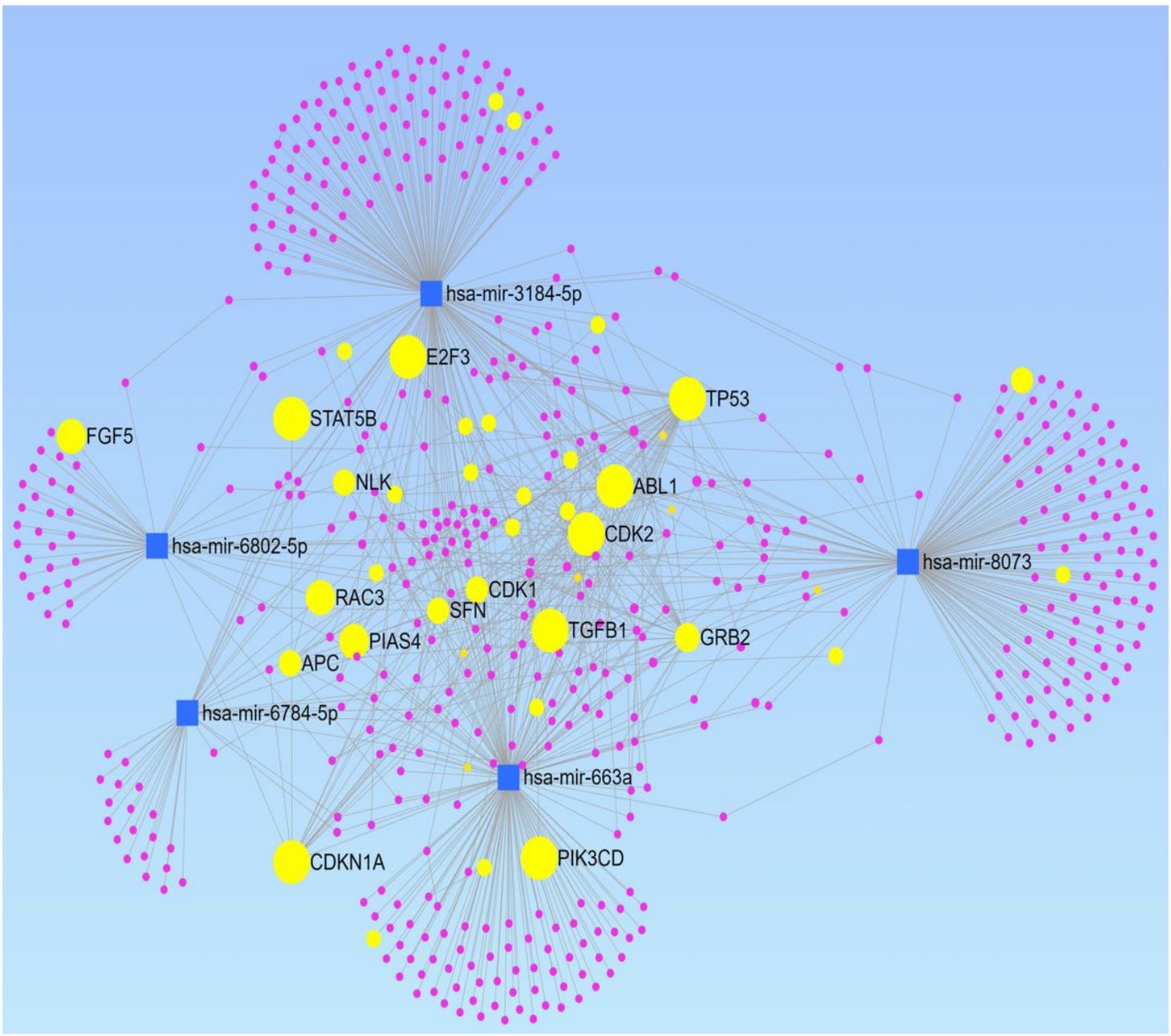

Figure 4

miRNA-mRNA Interaction Network for the selected 5 miRNAs. The blue squares represent the miRNAs. The blue and yellow circles represent the mRNAs. The yellow circles represent mRNAs directly associated with cancer, with the bigger yellow circles indicating that the mRNA is more associated with the selected 5 miRNAs. The edge between two nodes indicates their interaction.

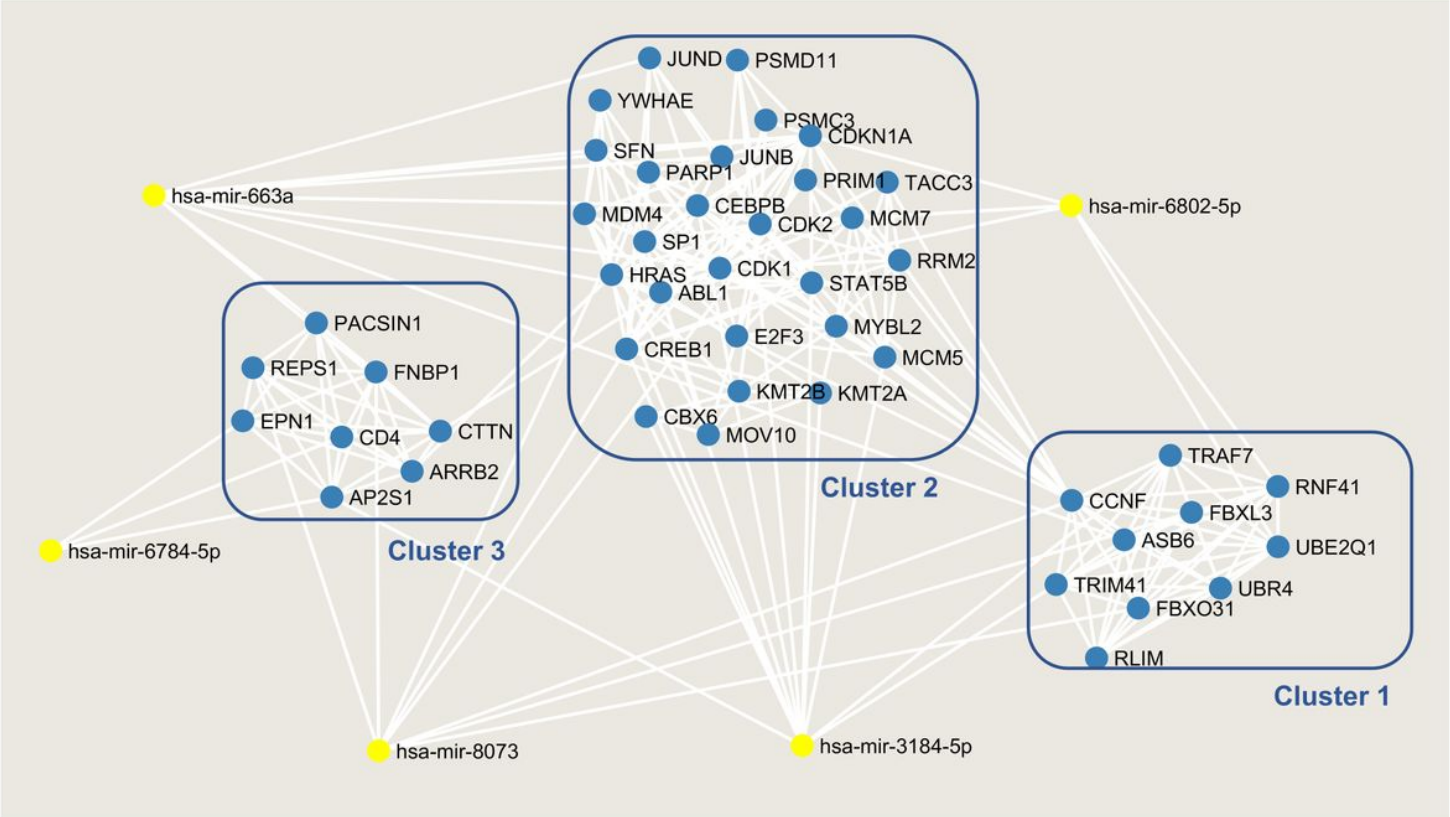

Figure 5

Cluster analysis of the mRNA presented in the miRNA-mRNA Interaction Network. Top clusters with MCODE value >5 from cytoscape were chosen and the clusters along with their interactions with the 5 selected miRNAs were shown. The miRNAs are highlighted, and the mRNAs are transparent.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx