SCIENTIFIC REPORTS

OPEN

Prioritizing cancer-related microRNAs by integrating microRNA and mRNA datasets

Daeyong Jin & Hyunju Lee

MicroRNAs (miRNAs) are small non-coding RNAs regulating the expression of target genes, and they are involved in cancer initiation and progression. Even though many cancer-related miRNAs were identified, their functional impact may vary, depending on their effects on the regulation of other miRNAs and genes. In this study, we propose a novel method for the prioritization of candidate cancer-related miRNAs that may affect the expression of other miRNAs and genes across the entire biological network. For this, we propose three important features: the average expression of a miRNA in multiple cancer samples, the average of the absolute correlation values between the expression of a miRNA and expression of all genes, and the number of predicted miRNA target genes. These three features were integrated using order statistics. By applying the proposed approach to four cancer types, glioblastoma, ovarian cancer, prostate cancer, and breast cancer, we prioritized candidate cancer-related miRNAs and determined their functional roles in cancer-related pathways. The proposed approach can be used to identify miRNAs that play crucial roles in driving cancer development, and the elucidation of novel potential therapeutic targets for cancer treatment.

MicroRNAs (miRNAs) are small non-coding RNAs that regulate the expression of target genes by binding to their 3′ untranslated regions. Recent studies aimed at the identification of cancer-related miRNAs revealed that miRNAs significantly affect cancer development by regulating the expression of oncogenes, tumor suppressors, and a large number of other genes, which results in the perturbation of biological networks[1,2]. Many computational approaches have been developed for the systemic identification of cancer-related miRNAs and their target genes and elucidation of the functional roles of miRNAs in cancer. These approaches can be broadly summarized into five categories.

First, several algorithms predict miRNA target genes based on the sequence complementary between these genes and miRNAs in the seed regions, and the predicted gene-miRNA interactions can be accessed through databases such as microCosm[3], Pictar[4], and TargetScans[5]. However, these predictions, based on sequences alone, cannot explain miRNA mechanisms in cancer development and progression, unless various biological activities, including miRNA-regulated gene and protein expression changes, are not considered.

Additionally, several computational approaches for the prediction of novel miRNA-disease relationships based on the existing biological databases, such as those containing information about miRNA similarities, disease similarities, and experimentally validated miRNA-disease relationships, have been proposed. Xuan et al.[6] assumed that miRNAs related to similar diseases are functionally related. Therefore, they calculated functional similarities between miRNAs based on previously known miRNA-disease relationships, and for each miRNA, they obtained k most similar miRNAs. These similarities were used to infer new miRNAs related to a given disease. Chen et al.[7] formulated miRNA-disease relationship prediction problem as an optimization problem based on regularized least squares using the same assumption as Xuan et al.[6], and demonstrated that the proposed approach successfully recovered miRNAs previously known to be related to several cancer types. In addition to this, they constructed miRNA-disease networks by employing Gaussian interaction profile kernel similarity[8,9] and restricted Boltzmann machines[10]. Pasquier et al.[11] used a vector space model to predict miRNA-disease relationships. They first combined miRNA-disease, miRNA-neighbor, miRNA-gene, miRNA-word, and miRNA-family relationships. Afterward, singular value decomposition was applied for dimension reduction, and miRNAs and disease were represented as vectors, and miRNAs related to diseases were prioritized based on vector similarities. Although

School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, 500-712, Korea. Correspondence and requests for materials should be addressed to H.L. (email: Hyunjulee@gist.ac.kr)

these approaches may uncover novel miRNA-disease relationships, they are highly dependent on the previously obtained knowledge, while predicting miRNAs with unknown relationships to any disease is difficult.

Next, several studies used miRNA expression changes to identify cancer-related miRNAs. Iorio et al.[12] showed that the differential expression of miRNAs in ovarian cancer can distinguish cancer cells from normal cells, and that over- and under-expression of miRNAs are associated with pathologic properties, such as histotype and lymphovascular invasion. Srinivasan et al.[13] identified ten signature miRNAs in glioblastoma (GBM) by analyzing miRNA expression data using COX regression analysis, and classifying patients into a low-risk and a high-risk group, based on the survival time. Zhang et al.[14] identified seven differentially expressed miRNAs (DE miRNAs), with their expression significantly associated with the survival time in hepatocellular carcinoma. They associated these seven signature miRNAs with several clinical parameters, such as tumor stage, tumor status, and gender, and found independent prognostic parameters based on univariate and multivariate analysis. DE miRNAs were identified, and survival analysis was performed in these studies in order to identify cancer-related miRNAs. However, some of these miRNAs may not be identified using this approach, due to several reasons, including the use of different processes for the filtering of clinical data, and different cohort sizes[15]. Additionally, miRNAs perform their functions in combination with transcription factors and other genes, and these combinatorial effects, and not miRNA expression alone, may be related to the survival[16].

Furthermore, computational approaches incorporating negative correlations between gene and miRNA expression levels have accelerated the identification of cancer-related miRNAs. MiRNA-gene pairs have been predicted based on various models, including linear regression, lasso regression, and Bayesian model, and these models have been applied to several cancer datasets[17]. The recent availability of paired miRNA and gene expression levels in multiple cancer datasets found in The Cancer Genome Atlas (TCGA)[18–21] allowed simultaneous analysis of miRNA and gene expression in multiple cancer types[22].

Finally, module-based approaches have been recently proposed, and here modules that contain a set of genes and miRNAs that are highly correlated and involved in the same pathways are identified by integrating multiple types of genomic data, such as gene and miRNA expression levels, gene-gene interactions, and gene-miRNA interactions[23–25]. These studies have highlighted the complex interactions between genes and miRNAs that contribute to the cancer development. Zhang et al.[23] employed a non-negative matrix factorization framework, where miRNA and gene expressions were factorized into a common basis matrix, and gene-miRNA regulatory modules were constructed. They showed that genes and miRNAs in these modules significantly overlaps with previously known cancer-related genes, miRNAs, and biological pathways. Zhang et al.[24] constructed a gene-miRNA network using negative correlations between gene and miRNA expressions and gene-miRNA interactions, and candidate cancer-related miRNAs were prioritized. Using our previous approach[25], we constructed gene-miRNA modules using a biclustering algorithm and a Gaussian Bayesian network framework. However, these approaches only consider local gene expression changes depending on miRNA expressions, but they do not consider the effects on the whole biological network.

Although significant efforts have been undertaken to identify cancer-related miRNAs, only a small number of studies prioritizes cancer-related miRNAs based on gene and miRNA expressions. For mRNAs, several algorithms have been developed in order to prioritize disease-related genes[26,27], because the functional effects of genes may differ. Similarly, the extents of functional effects of miRNAs in cancer may vary depending on how miRNA expression changes are propagated through the biological network.

Here, we aimed to prioritize miRNAs that lead to significant changes in the whole biological network during cancer initiation and development. We propose a novel approach based on order statistics that prioritizes miRNAs whose expression changes significantly affect cancer development. Additionally, we explain functional roles of the miRNAs highly ranked in our model at the pathway level.

Methods

As shown in Fig. 1, we assumed that some genes and miRNAs, illustrated in the left part of the presented biological network, have many interactions and cause significant alterations of other genes/miRNAs, directly or indirectly. Therefore, they are more likely to contribute to the development of cancer than genes and miRNAs represented in the right part of the network, which are mainly affected by the previous category of genes and miRNAs, and are less likely to contribute to the cancer development. A schematic overview of our approach is presented in Fig. 2, while the details are described below.

Data collection. We obtained microarray and/or RNA-Seq datasets for GBM, ovarian cancer (OVC), prostate cancer (PRCA), and breast cancer (BRCA) from the TCGA data portal (http://cancergenome.nih.gov). Paired datasets of gene and miRNA expressions were obtained.

For microarray, gene and miRNA expression data were generated using Affymetrix HG-U133A and Agilent H-miRNA_8 × 15 for GBM and OVC, respectively. GBM dataset contains 12,042 genes and 470 mature miRNAs, obtained from 482 tumor samples, and OVC dataset contains 12,042 genes and 723 mature miRNAs obtained from 578 tumor samples. For RNA-Seq, gene and miRNA expression datasets were generated by IlluminaHiSeq_RNASeqV2 and IlluminaHiSeq_miRNASeq, respectively, using OVC, PRCA and BRCA samples. OVC dataset contains 20,806 genes from 416 tumor samples, PRCA dataset contains 20,531 genes from 494 tumor samples, and BRCA dataset contains 20,532 genes from 461 tumor samples. Additionally, they commonly contain 1,046 miRNAs obtained from the paired samples with genes.

Predicted gene-miRNA interactions were collected from microCosms[3], PicTar[4], and TargetScans[5]. The information about miRNA-disease relationships was obtained from the Human microRNA Disease Database (HMDD)[28]. We collected OVC miRNA data using "Ovarian Neoplasm" term, GBM miRNAs using "Glioblastoma" or "Glioma" terms, PRCA miRNA data using "Prostatic Neoplasms" term, and BRCA miRNA data using "Breast Neoplasms" term.
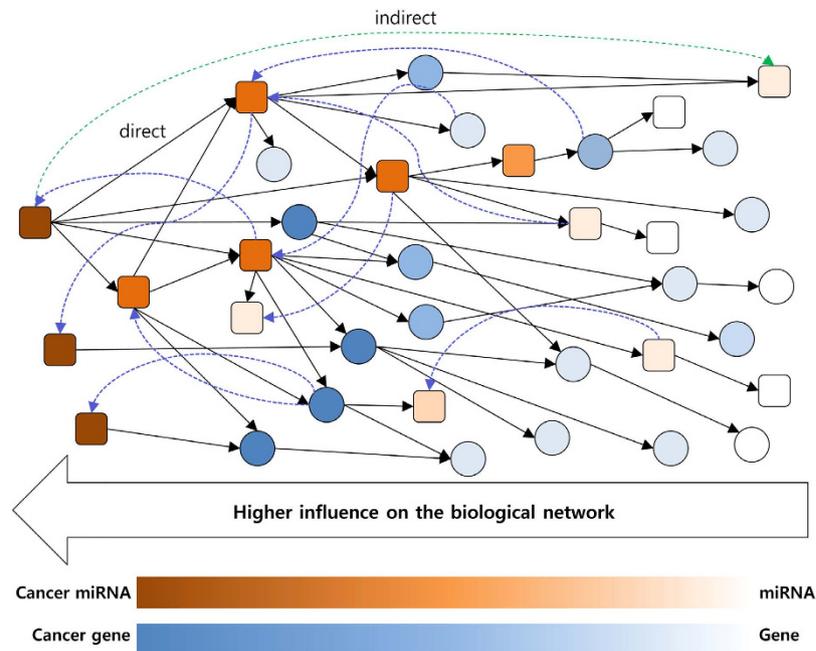
Figure 1. Biological network structure. Cancer-related miRNAs (dark orange rectangles) and genes (dark blue circles) affect other genes and miRNAs and ultimately the entire biological network. In contrast to this, white rectangles and circles represent miRNAs and genes that do not affect other genes and these miRNAs are mostly regulated by the previous category of genes and miRNAs. Black solid lines and green dotted lines represent the direct or indirect changes, respectively, of one node that affects linked nodes. Although we presented only one indirect edge for simplicity, there are many indirect relationships in a real biological network. Right nodes influence on the left nodes shown in blue dotted lines, however, influences from left to right are more abundant globally.

Feature selection and analysis. We propose three miRNA features that may strongly affect the biological network. First, we assumed that highly expressed miRNAs show a high potential to affect the biological network. Therefore, the average miRNA expressions from all cancer samples are selected as the first feature (F1). $F_a = \{a_1, \ldots, a_i, \ldots, a_M\}$, where a_i is the average expression of a miRNA i in the cancer samples, and M represents the number of miRNAs. Note that we only considered the expression levels of miRNAs in cancer cells, but not in the normal cells.

Additionally, we assumed that if a miRNA significantly affects some genes, the expressions of this miRNA and the genes may be highly correlated. A miRNA can directly regulate a set of genes, which may indirectly lead to the alterations in the expression of many other genes. Therefore, we considered all genes in the biological network and used the average of absolute Pearson correlation coefficients (PCCs) between miRNA and all gene expressions as the second feature (F2). $F_c = \{c_1, \ldots, c_i, \ldots c_M\}$, where c_i is the average of absolute PCC values between miRNA i and all genes in the cancer samples.

We further assumed that miRNAs that bind to many genes strongly affect the biological network. However, only a small fraction of miRNA target genes has been experimentally validated, and therefore, we used computationally predicted gene-miRNA interactions, based on sequence complementary. We obtained the predicted gene-miRNA interactions from microCosms[3], PicTar[4], and TargetScans[5]. All interaction pairs were extracted from these three databases and duplicated interaction pairs were removed. Furthermore, we counted the number of the predicted targets for each miRNA and these numbers can be considered the numbers of potential interacting genes, representing our third feature (F3). $F_t = \{t_1, \ldots, t_i, \ldots, t_M\}$, where t_i is the number of target genes for miRNA i. Note that, because the number of predicted targets is determined by the sequence complementary information, gene-miRNA interactions for the third feature are the same, regardless of the cancer type used.

Integration of features. We used order statistics to integrate the three features, F1, F2, and F3. In Fig. 3, a flowchart of feature integration process is presented. First, we computed the ranking ratios for the values of each feature. F_a, F_c, and F_t values were ranked in a decreasing order and their ranking ratios were stored in RD_1, RD_2, and RD_3, respectively. Let $RD_1 = \{ad_1^r, \ldots, ad_i^r, \ldots, ad_j^r, \ldots, ad_M^r\}$, where ad_i^r is the ranking ratio of $a_i \in F_a$ and $ad_i^r < ad_j^r$ if $a_i > a_j$. Similarly, $RD_2 = \{cd_1^r, \ldots, cd_i^r, \ldots, cd_j^r, \ldots, cd_M^r\}$, where cd_i^r is the ranking ratio of $c_i \in F_c$ and $cd_i^r < cd_j^r$ if $c_i > c_j$, and $RD_3 = \{td_1^r, \ldots, td_i^r, \ldots, td_j^r, \ldots, td_M^r\}$, where td_i^r is the ranking ratio of $t_i \in F_t$ and $td_i^r < td_j^r$ if $t_i > t_j$. Furthermore, for a given miRNA i, a Q statistic is computed based on joint cumulative distribution of order statistic of the three features[26]. The integration was approximated using a recursive solution proposed by Stuart *et al.*[29].
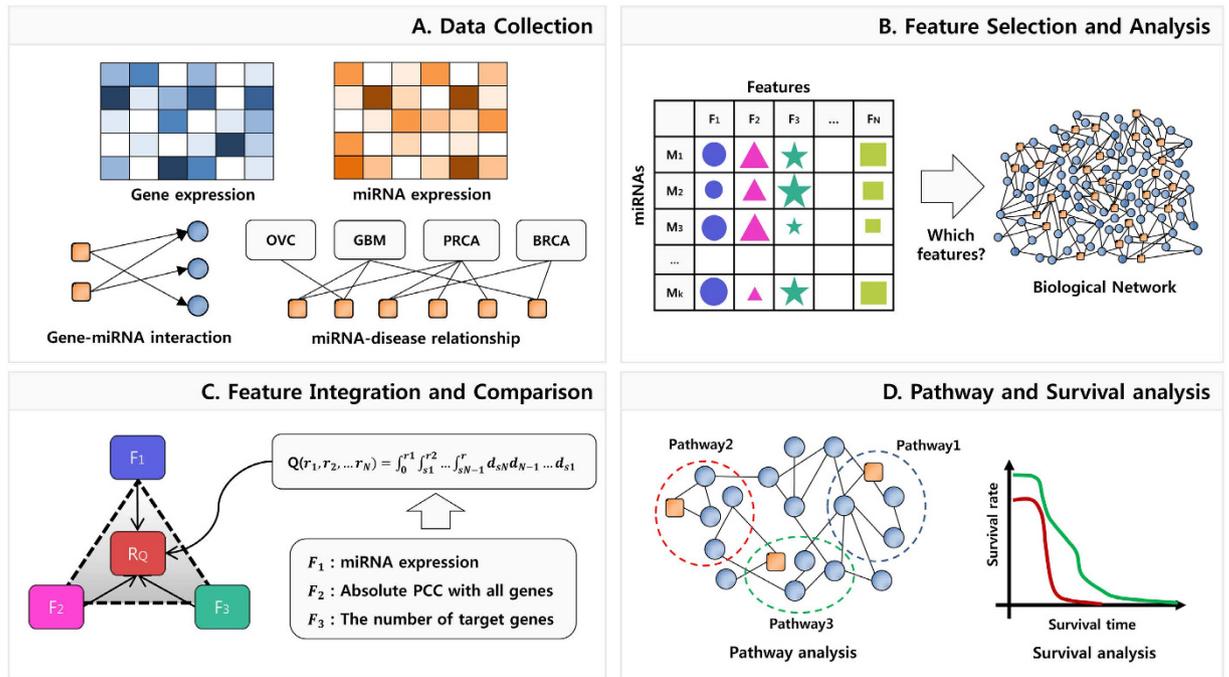
Figure 2. Schematic overview of the proposed approach. (**A**) Gene and miRNA expressions in paired samples, gene-miRNA interactions, and miRNA-disease relationships are collected. (**B**) Selection and analysis of cancer-related miRNA features. (**C**) Integration of features F1, F2, and F3 selected in the step (**B**) and calculation of integrated rankings R_Q using order statistic. (**D**) Pathway and survival analysis for the understanding of functional roles of miRNAs in biological pathways.
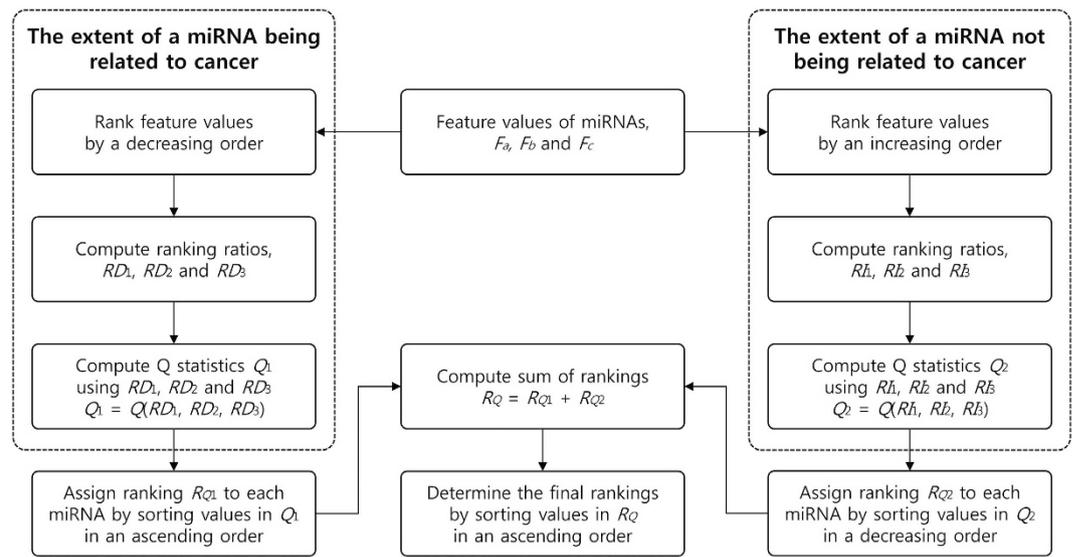

Figure 3. Flowchart showing the feature integration process. We computed feature values for each miRNA. Following this, we computed ranking ratios by decreasing (*RD*) and increasing order (*RI*). Afterward, we computed Q statistics, Q_1 and Q_2, and rankings, R_{Q1} and R_{Q2}. Finally, we determined the final rankings by the addition of these two rankings.

$$Q(r_1, r_2, \ldots, r_N) = N! \int_0^{r_1} \int_{s_1}^{r_1} \ldots \int_{s_{N-1}}^{s_N} ds_N ds_{N-1} \ldots ds_1$$

$$Q(r_1, r_2, \ldots, r_N) = N! \sum_{k=1}^{N} (r_{N-k+1} - r_{N-k}) \, Q(r_1, r_2, \ldots, r_{N-k}, r_{N-k+2}, \ldots, r_N),$$

where $N = 3$, the number of the features, $r_0 = 0$, and r_k is the ranking ratio for the kth feature for the given miRNA i: $r_1 = ad_i^r \in RD_1$, $r_2 = cd_i^r \in RD_2$, and $rd_3 = td_i^r \in RD_3$. For all miRNAs, Q statistics were calculated, so that $Q_1 = \{qd_1, ..., qdi, ..., qd_M\}$, where qd_i is the Q statistic for a miRNA i. More specifically, $qd_i = Q(ad_i^r, cd_i^r, td_i^r)$.

We assigned a ranking to each miRNA by sorting values in Q_1 in an ascending order. Let $R_{Q1} = \{qd_1^r, ..., qd_i^r, ..., qd_j^r, ..., qd_M^r\}$, where qd_i^r is a ranking of $qd_i \in Q_1$ and $qd_i^r < qd_j^r$ if $qd_i < qd_j$. We assumed that a miRNA i with a smaller value of qd_i (thus, with a higher ranking) is more highly related to cancer development, since we assumed that miRNAs with larger F_a, F_c, and F_t values are more likely to be related to cancer, resulting in smaller qd_i values. Therefore, R_{Q1} is a set of miRNA rankings, indicating their relevance to cancer development. However, some miRNAs are highly ranked in some features but not in others. In these cases, miRNA rankings can be dominantly determined by a single or a small number of features because the ranking in R_{Q1} becomes higher with a lower number of high ranking features. To reduce the effects produced by these few features, we incorporated the following step.

We calculated another Q statistic that measures the extent to which a miRNA is not related to cancer. When F_a, F_c, and F_t feature values are small, it is unlikely that a miRNA is related to cancer. Hence, F_a, F_c, and F_t values are ranked in an ascending order and their ranking ratios are stored in RI_1, RI_2, and RI_3, respectively. Let $RI_1 = \{ai_1^r, ..., ai_i^r, ..., ai_j^r, ..., ai_M^r\}$, where ai_i^r is the ranking ratio of $a_i \in F_a$ and $ai_i^r > ai_j^r$ if $a_i > a_j$. Similarly, $RI_2 = \{ci_1^r, ..., ci_i^r, ..., ci_j^r, ..., ci_M^r\}$, where ci_i^r is the ranking ratio of $c_i \in Fc$ and $ci_i^r > ci_j^r$ if $c_i > c_j$, and $RI_3 = \{ti_1^r, ..., ti_i^r, ..., ti_j^r, ..., ti_M^r\}$, where ti_i^r is the ranking ratio of $t_i \in F_t$ and $ti_i^r > ti_j^r$ if $t_i > t_j$. Afterward, rankings in RI_1, RI_2, and RI_3 are integrated using Q statistics. Let $Q_2 = \{qi_1, ..., qi_i, ..., qi_M\}$, where qi_i is the Q statistic for a miRNA i obtained by integrating its rankings in RI_1, RI_2, and RI_3. More specifically, $qi_i = Q(ai_i^r, ci_i^r, ti_i^r)$. Because Q_2 measures the extent of a miRNA not being related to cancer, we sorted them in a decreasing order and stored the rankings into R_{Q2}. Let $R_{Q2} = \{qi_1^r, ..., qi_i^r, ..., qi_j^r, ..., qi_M^r\}$, where qi_i^r is a ranking of $qi_i \in Q_2$ and $qi_i^r > qi_j^r$ if $qi_i < qi_{ij}$. R_{Q2} represents a set of miRNA rankings, indicating their relevance to cancer development.

Final rankings were determined using a sum of these two rankings: $R_Q = R_{Q1} + R_{Q2}$. We used R_Q as a new statistic to determine all miRNA rankings, penalizing miRNA rankings that did not show significant feature values for some of the three examined features. Several examples of integrating features are described in Supplementary Table 1.

Pathway analysis related to miRNA.   In order to understand the functional roles of miRNAs in biological pathways, we performed a pathway enrichment analysis using KEGG pathways and gene ontology (GO) biological functions. We downloaded 186 KEGG pathways and 825 GO biological functions from Gene Set Enrichment Analysis (GSEA, http://www.broadinstitute.org/gsea), and extracted experimentally validated gene-miRNA interaction datasets from miRTarbase[30]. Strong evidence, such as reporter assays, western blotting, and qPCR and/or weak evidence, such as microarrays, next generation sequencing (NGS), and pSILAC support these datasets. For each miRNA, we selected genes with expression levels highly correlated with the given miRNA expression levels (1% of genes in PCC values) or those interacting with that miRNA according to miRTarbase[30]. Finally, we performed a pathway enrichment analysis using a hypergeometric test and obtained p-values for each miRNA. p-values were corrected to q-values based on Benjamini & Hochberg correction[31], in order to address multiple comparison issues. We considered pathways with q-value less than 0.05 as significant.

Survival analysis.   We performed a survival analysis to identify miRNAs that play crucial roles in cancer patient survival. Clinical information about GBM and OVC samples was obtained from TCGA. For each candidate miRNA, we divided these samples into two groups: an under-expressed group, where the expression levels of that miRNA belong to the bottom 10% of all values, and an over-expressed group, where the expression levels of the miRNA belong to the top 10% of values. We performed Kaplan-Meier survival analysis and obtained p-values.

Results

Feature selection and analysis.   For each miRNA, we computed three feature values, F1, F2, and F3. To show that these three features are cancer-related, we ranked all miRNAs based on the feature values. We assigned a value of one to miRNAs if they were found in the HMDD database and zero if not. We computed a cumulative ratio by computing the average number of miRNAs that had the value of one for each ranking. Afterward, we determined whether the higher-ranked miRNAs have higher cumulative ratios. In Fig. 4, cumulative ratios of cancer-related miRNAs for the five cancer datasets are presented. Images on the left in Fig. 4 show that cumulative ratios of all features, F1, F2, and F3, have a similar tendency of steady decrease with the ranking, suggesting that highly ranked miRNAs are more likely to be related to cancer regardless of cancer type (GBM, OVC, PRCA, and BRCA) and platforms (microarray or RNA-Seq).

Integration of features.   We integrated the three features (F1, F2, and F3) using order statistics and ranked miRNAs based on R_Q. As shown in Fig. 4, the cumulative ratios of cancer-related miRNAs were shown to be the highest when these three features were integrated, compared with single-feature or two-feature analyses. Additionally, we applied different integration methods: an average of ranking ratios of the three features, F1, F2, and F3 (e.g. $(RD_1 + RD_2 + RD_3)/3$), an inverse normal transformation[32], and two other order statistics, R_{Q1} and R_{Q2}. The comparison of the proposed integration method (R_Q) with the other tested methods showed that our method had the highest performance on average when GBM, OVC (microarray), OVC (RNA-Seq), PRCA, and BRCA samples were used (Supplementary Table 2). These results suggest that the integration of these features assists the identification of cancer-related miRNAs. Supplementary Tables 3, 4, 5, 6, and 7 show miRNA rankings, integrated scores and cumulative ratios of cancer-related miRNAs for GBM, OVC (microarray), OVC (RNA-Seq), PRCA, and BRCA samples, respectively.

A. Glioblastoma
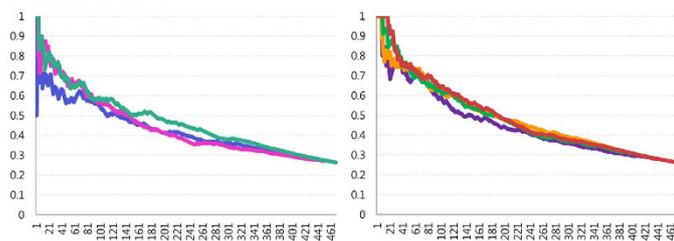
Rank	F1	F2	F3	F1+F2	F1+F3	F2+F3	F1+F2+F3
≤ 10	0.70	0.80	0.90	0.80	0.90	1.00	1.00
≤ 20	0.65	0.85	0.80	0.75	0.80	0.85	0.95
≤ 30	0.63	0.73	0.77	0.80	0.77	0.83	0.83
≤ 50	0.58	0.66	0.68	0.76	0.74	0.74	0.74
≤ 100	0.55	0.57	0.59	0.58	0.62	0.63	0.64
≤ 200	0.42	0.42	0.48	0.45	0.48	0.48	0.48
≤ 300	0.36	0.34	0.38	0.36	0.39	0.37	0.37
≤ 400	0.31	0.30	0.31	0.30	0.31	0.31	0.31
≤ 470	0.26	0.26	0.26	0.26	0.26	0.26	0.26

B. Ovarian Cancer (microarray)
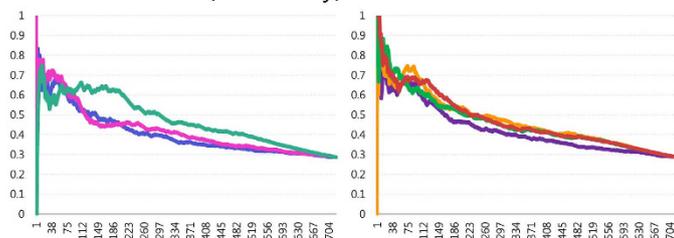
Rank	F1	F2	F3	F1+F2	F1+F3	F2+F3	F1+F2+F3
≤ 10	0.80	0.70	0.70	0.80	0.80	0.80	0.90
≤ 20	0.70	0.70	0.70	0.70	0.70	0.80	0.80
≤ 30	0.63	0.70	0.60	0.63	0.70	0.80	0.73
≤ 50	0.68	0.68	0.56	0.60	0.66	0.72	0.66
≤ 100	0.54	0.58	0.63	0.62	0.69	0.64	0.67
≤ 200	0.46	0.46	0.62	0.47	0.54	0.52	0.53
≤ 300	0.40	0.42	0.50	0.41	0.47	0.49	0.45
≤ 500	0.33	0.34	0.40	0.34	0.39	0.45	0.38
≤ 723	0.29	0.29	0.29	0.29	0.29	0.29	0.29

C. Ovarian Cancer (RNA-Seq)
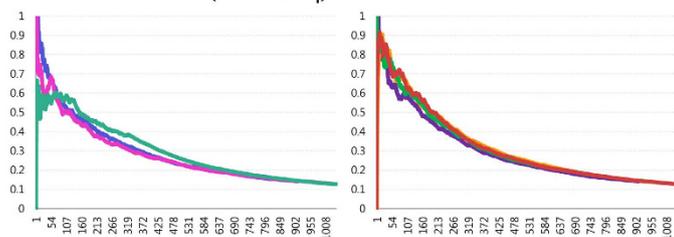
Rank	F1	F2	F3	F1+F2	F1+F3	F2+F3	F1+F2+F3
≤ 10	0.90	0.70	0.60	0.80	0.80	0.90	0.90
≤ 20	0.85	0.70	0.55	0.80	0.90	0.85	0.85
≤ 30	0.77	0.60	0.57	0.77	0.83	0.73	0.83
≤ 50	0.68	0.66	0.58	0.66	0.76	0.76	0.74
≤ 100	0.53	0.51	0.56	0.59	0.66	0.60	0.66
≤ 200	0.42	0.39	0.46	0.44	0.48	0.47	0.50
≤ 300	0.34	0.32	0.39	0.35	0.38	0.37	0.37
≤ 500	0.23	0.23	0.26	0.23	0.26	0.24	0.25
≤ 1046	0.13	0.13	0.13	0.13	0.13	0.13	0.13

D. Prostate Cancer
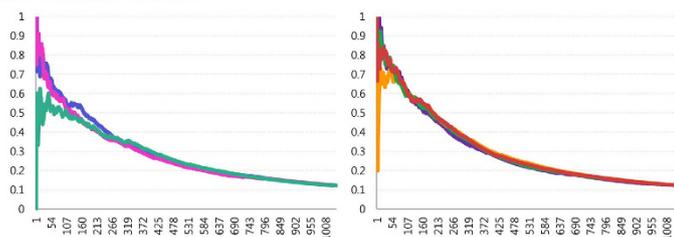
Rank	F1	F2	F3	F1+F2	F1+F3	F2+F3	F1+F2+F3
≤ 10	0.78	0.89	0.33	0.90	0.60	0.90	0.80
≤ 20	0.75	0.85	0.55	0.80	0.65	0.86	0.80
≤ 30	0.73	0.70	0.50	0.80	0.63	0.80	0.80
≤ 50	0.66	0.66	0.54	0.74	0.72	0.74	0.76
≤ 100	0.56	0.55	0.49	0.62	0.60	0.60	0.63
≤ 200	0.47	0.42	0.43	0.47	0.50	0.48	0.50
≤ 300	0.34	0.34	0.35	0.35	0.38	0.37	0.38
≤ 500	0.23	0.23	0.24	0.23	0.25	0.23	0.24
≤ 1046	0.12	0.12	0.12	0.12	0.12	0.12	0.12

E. Breast Cancer
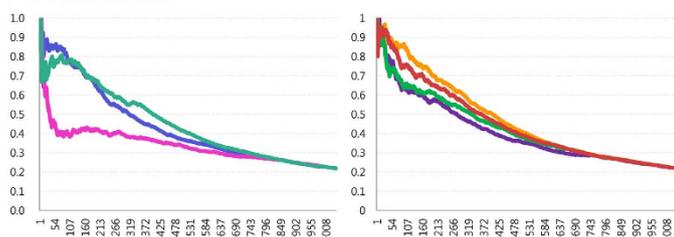
Rank	F1	F2	F3	F1+F2	F1+F3	F2+F3	F1+F2+F3
≤ 10	0.90	0.80	0.70	1.00	0.90	0.90	0.90
≤ 20	0.85	0.60	0.75	0.95	0.95	0.90	0.95
≤ 30	0.87	0.53	0.70	0.87	0.97	0.83	0.93
≤ 50	0.84	0.46	0.76	0.76	0.88	0.72	0.86
≤ 100	0.78	0.40	0.77	0.63	0.85	0.65	0.75
≤ 200	0.65	0.42	0.67	0.58	0.70	0.61	0.65
≤ 300	0.52	0.39	0.56	0.48	0.61	0.52	0.57
≤ 500	0.37	0.34	0.42	0.36	0.41	0.38	0.39
≤ 1046	0.22	0.22	0.22	0.22	0.22	0.22	0.22

Legend: F1 — F2 — F3 — F1+F2 — F1+F3 — F2+F3 — F1+F2+F3

Figure 4. Performances of features selected for the ranking of cancer-related miRNAs. *x*-axis represents miRNA rankings and *y*-axis represents a cumulative ratio of cancer-related miRNAs included in HMDD. (**A–E**) correspond to GBM, OVC (microarray), OVC (RNA-Seq), PRCA and BRCA samples, respectively. Left, cumulative ratios of cancer-related miRNAs are presented, which were obtained using a single feature. Right, cumulative ratios of cancer-related miRNAs, by integrating features, are presented. For most cancer types, cumulative ratios of cancer-related miRNAs increase with the rank. Additionally, the integration of more than one feature shows higher performance than when only one feature is used. Furthermore, integration of all features shows the highest performance for most cancer types.

GBM-related miRNAs. Of 470 miRNAs found in the GBM dataset, 123 miRNAs (26.3%) were shown to be related to GBM according to HMDD. In our analysis, 19 out of the top 20 miRNAs (95.0%) and 64 out of the top

100 miRNAs (64.0%) were shown to be related to GBM (Fig. 4A), which demonstrates a significant overlap with GBM-related miRNAs (p-value = 4.43e−20, in the hypergeometric test).

Furthermore, we investigated whether 98 previously known GBM-related genes[33,34] interact with miRNAs ranked high in our analysis. The target genes of the top 20 ranked miRNAs obtained from miRTarbase[30] included 3949 genes and 515 genes, found to interact with at least one of the miRNAs with some evidence and with strong evidence, respectively. Among these, 68 and 43 genes were common with the 98 GBM genes, with some evidence and strong evidence, respectively, showing statistically significant overlaps (p-values < 1.12e−13 in the hypergeometric test). Extracted gene-miRNA interactions are listed in Supplementary Table 8.

OVC-related miRNAs–microarray dataset.   Of 723 miRNAs in OVC microarray dataset, 206 (28.6%) were found in HMDD. In our analysis, 16/20 miRNAs (80%) and 67/100 (67.0%) miRNAs were shown to be OVC-related miRNAs included in HMDD (Fig. 4B), showing a significant overlap (p-value = 3.41e-18 in the hypergeometric test).

We further investigated whether previously known 379 OVC genes[33,34] interact with the highly ranked miRNAs. The top 20 ranked miRNAs are known to interact with 4485 genes and 559 genes with some evidence and strong evidence in miRTarbase[30], respectively. Of these genes, 183 and 89 genes were found among 379 OVC genes with some evidence and strong evidence, respectively, showing statistically significant overlaps (p-values < 5.37e-06 in the hypergeometric test). In Supplementary Table 9, a list of gene-miRNA interactions is presented.

OVC-related miRNAs–RNA-Seq dataset.   In RNA-Seq OVC dataset, among 1046 identified miRNAs, 133 (12.7%) are included in HMDD as OVC-related. Our analyses showed that 17/20 miRNAs (85%) and 66/100 (66.0%) are included in HMDD as OVC-related genes (Fig. 4C), demonstrating a significant overlap with OVC-related miRNAs (p-value = 8.88e-42 in the hypergeometric test).

PRCA and BRCA-related miRNAs.   PRCA dataset contained 1046 miRNAs, and 128 (12.2%) are indicated in HMDD as well. In our analysis, 16 out of 20 miRNAs (80.0%) and 63 out of 100 miRNAs (63.0%) were shown to be PRCA-related (Fig. 4D). In the BRCA dataset, 229 (21.9%) miRNAs, of 1046, were found in HMDD as well. Our analysis, integrating two features, F1 and F3, showed the best performance in the identification of BRCA-related miRNAs. However, the integration of all features shows good performance in general. This analysis identified 19 out of 20 miRNAs (95.0%) and 75 out of 100 miRNAs (75.0%) as related to BRCA (Fig. 4E).

Pathway analysis.   We investigated biological pathways affected by cancer-related miRNAs identified in our study. For each miRNA, we calculated PCCs between the expression level of all genes and the given miRNA, and selected highly correlated genes, with PCC values within top 1%. These highly correlated genes and genes that are reported to interact with a specific miRNA in miRTarbase[30] are considered candidate genes related to the given miRNA.

Furthermore, we constructed reference cancer-related pathways for GBM and OVC. We performed a pathway enrichment test using 98 GBM genes curated from two previous reports[33,34], and 379 OVC genes curated from Dragon Database For Exploration Of Ovarian Cancer Genes (DDOC)[35] against KEGG pathways. As a result, 61 and 73 pathways were shown to be significantly enriched (q-value < 0.05) and used as reference GBM and OVC, respectively, related pathways (Supplementary Tables 10 and 11).

Pathway enrichment analysis, using the genes related to top 20 GBM candidate miRNAs, is presented in Fig. 5A. A number of pathways (81) were shown to be associated with top 20 miRNAs (Supplementary Table 12). Among them, 47 pathways are reference cancer-related pathways. They include known GBM-related pathways, such as apoptosis[36], cell cycle[37], cytokine-cytokine receptor interaction[38], ErbB signaling pathway[39], JAK-STAT signaling pathway[40], MAPK and mTOR signaling pathway[41], p53 signaling pathway[42], TGF-beta signaling pathway[43], VEGF signaling pathway[44], and Wnt signaling pathway[45]. Additionally, several associations between miRNAs and GBM-related pathways shown in Fig. 5A were previously reported. MiR-181a, miR-21, miR-34a, miR-30a-5p, miR-222, and miR-17-5p were shown to be involved in the apoptosis of glioma cells[46–51]. MiR-21, miR 34a, and miR-30a-5p may be involved in ErbB signaling pathway[52–54] and miR-21, miR-34a, miR-125b, miR-17-5p/3p, miR-106a, and miR-222 were identified in the p53 signaling pathway in GBM[50,51,55–58]. MiR-30a is involved in Wnt signaling pathway through the regulation of PRDM1 during glioma cell growth[59]. Kwak et al.[60] showed that miR-21 is involved in glioma invasion by controlling MAPK signaling pathway.

In OVC analysis, 84 pathways were associated with the top 20 OVC candidate miRNAs (Fig. 5B and Supplementary Table 13). Among them, 56 pathways represent reference cancer-related pathways for OVC, including previously known OVC related pathways, such as apoptosis[61], ECM receptor interaction[62], ErbB signaling pathway[63], JAK-STAT signaling pathway[64], p53 signaling pathway[65], and TGF-β signaling pathway[66]. Additionally, in Fig. 5B, previously known associations between miRNAs and pathways are presented. MiR-21 and miR-17 were shown to be involved in OVC cell apoptosis[67,68] in p53 signaling pathway in OVC cells[69,70].

Additionally, we performed a functional enrichment test for GO biological functions using the genes related to top 20 GBM and OVC candidate miRNAs. Among 1123 enriched terms for GBM, 290 terms were related to cancer hallmark signatures such as apoptosis, immune response, chromosome abnormalities, inflammation, and angiogenesis[71,72]. Similarly, for OVC, 171 out of 791 enriched terms were related to cancer, showing that the top ranked miRNAs significantly affect cancer-related pathways. Enriched terms are listed in Supplementary Tables 14 and 15.

Survival analysis.   We obtained clinical information about 480 GBM and 561 OVC (microarray) samples. Out of 480 GBM patients, 377 were shown to be deceased and the average survival is 493 days. Out of 561 OVC
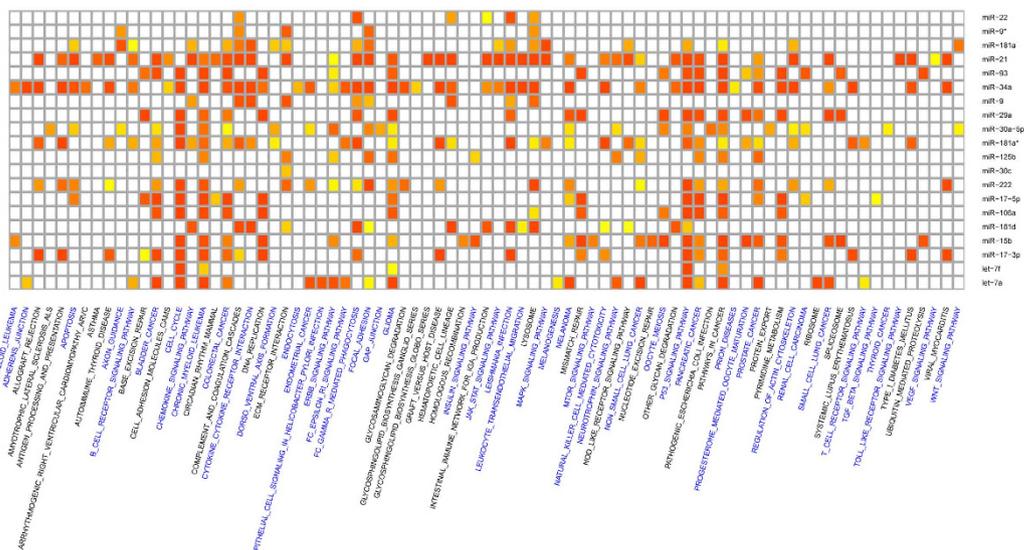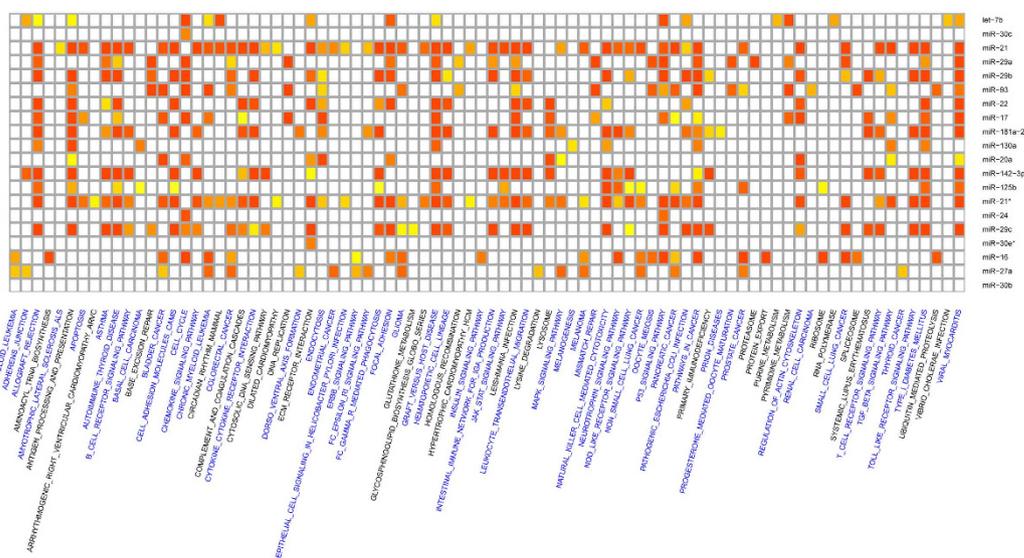
Figure 5. GBM and OVC (microarray) pathway analysis. (A,B) Show heatmaps of pathways related to the top 20 miRNAs obtained by analyzing GBM and OVC samples, respectively. Red, orange, yellow, and white colors represent the correlation of a miRNA and the pathway (form high to low, based on *p*-value using a hypergeometric test). Cancer-related pathways are presented in blue.

patients, 290 were deceased and the average survival is 1088 days. Considering these data, GBM is more aggressive than OVC. For each miRNA, we divided the samples into 10% underexpressed group and 10% overexpressed group and performed the survival analysis. For GBM samples, 10 out of the top 20 miRNAs (50%) and 29 out of the top 100 miRNAs (29%) were shown to be significantly related to the survival of patients. In Fig. 6, survival curves for samples containing overexpressed or underexpressed miR-17-5p, miR-106a, and miR-181d, which function as tumor suppressors, are shown. Overexpressed groups show longer survival time compared with that of the underexpressed groups. These miRNAs were previously identified as tumor suppressors, as presented in Table 1. For OVC, only six out of the top 100 miRNAs (6%) were shown to be related to the survival of patients. Although multiple studies showed that the highly ranked miRNAs identified in this study are related to OVC, for most of them, no significant difference in the survival time was observed. This may be a result of the higher survival ratio of patients with OVC compared with GBM patients and relatively small differences in the expression levels between the samples. Details of the survival analysis for the top 100 miRNA are described in Supplementary Tables 16 and 17 for GBM and OVC samples, respectively.

Comparison with other methods. We compared performances of other methods with that of our method using GBM and OVC (microarray) datasets. We used differentially expressed miRNAs for comparison, because
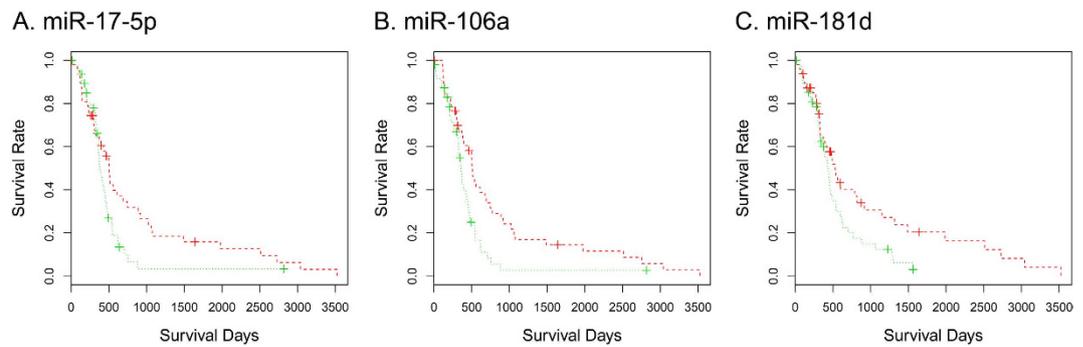
Figure 6. GBM survival analysis. Green and red lines represent samples with underexpressed miRNAs and overexpressed miR-17-5p, miR-106a, or miR-181d, respectively. These miRNAs function as tumor suppressors, and the survival rates of patients with the overexpression of these miRNAs are higher than those of patients with the underexpression of these miRNAs.

the identification of miRNAs based on differential expression between cancer samples and normal samples is the most commonly used approach for the selection of cancer-related miRNAs. We additionally downloaded eight and 10 unmatched normal GBM and OVC (microarray) samples, respectively, computed p-values using t-test, and ranked miRNAs based on p-values.

For comparison, we employed our previous approach[25] as well, in which gene and miRNA expressions and gene-gene interactions are integrated using a biclustering algorithm and a Gaussian Bayesian network framework, followed by construction of gene-miRNA modules. The inclusion of genes and miRNAs in the same module was explained through the direct regulation of genes by miRNAs and/or indirect regulation by transcription factors. We assumed here that the miRNAs included in the modules are cancer-related. We listed all miRNAs in the modules without duplication and obtained 95 GBM-related and 108 OVC-related candidate miRNAs.

In Fig. 7, each bar represents a cumulative ratio of GBM miRNAs in HMDD (Fig. 7A) and OVC miRNAs in HMDD (Fig. 7B) among miRNAs identified by each method, and on the x-axis are represented the numbers of candidate miRNAs selected based on the rankings in each method. Across all rankings ranging from 10 to 100, our method outperformed the other two investigated approaches. Although our method showed a slightly higher performance when 95 GBM candidate miRNAs were used, in comparison with our previous approach (66.3% vs. 61.1%, respectively), it significantly outperformed our previous approach when using 108 OVC candidate miRNAs (66.3% vs 48.1%, respectively).

An approach proposed by Zhang et al.[24] was also employed for comparison. Zhang et al.[24] identified differentially expressed genes and miRNAs, computed PCCs between them, and constructed a gene-miRNA network by intersecting negative PCC pairs and combination of experimentally and computationally derived gene-miRNA interactions. They ranked candidate cancer-related miRNAs by considering the number of genes uniquely regulated by miRNAs in the obtained network. When the method was applied to a GSE34933 dataset from NCBI GEO, containing the information about gene and miRNA expression for four PRCA samples and benign prostatic hyperplasia samples, 26, 39 and 69 PRCA candidate miRNAs were identified depending on the thresholds. In Fig. 7C, it is clearly demonstrated that our method outperforms the approach used by Zhang et al.[24] when the same PRCA dataset is used.

Discussion and Conclusion

Our approach can identify cancer-related miRNAs that may affect other molecules in the biological network based on three distinct features. For most of the analyzed cancer types, the integration of the three features led to the best performance in the identification of cancer-related miRNAs. However, the identification of BRCA-related miRNAs was improved when only two features, F1 and F3, were integrated, because F2 showed less power. Although the influence of these features can be controlled by assigning different weights to each feature, this requires prior knowledge, and therefore, the ability of generalization decreases. Here, we demonstrated that our order statistics-based method outperformed the average ranking ratio approach. To further investigate the influence of different weighting of features, we determined performance of feature integration by assigning different weights to each feature. These three features were integrated with various weights ranging from 0.1 to 1.0, resulting in a total of 1,000 test cases for each cancer type. We then selected the weight combinations providing the best performance, and the best combinations yielded performances similar to our approach for GBM, OVC and PRCA when selecting 100 candidate miRNAs, although our approach was outperformed in other cases. However, the best weight combinations were differed between different cancer types and the number of miRNAs selected as candidate cancer-related miRNA, confirming that it is hard to generalize optimal feature weights. Additionally, our order statistics approach significantly outperformed the worst weight combination, showing that our approach that considers ascending and descending rankings of the features can be generalized to any cancer type. Performances with various feature weights are described in Supplementary Table 18.

For GBM and OVC, we investigated top 20 miRNAs in detail to understand their roles in the cancer development, because more highly ranked miRNAs are more likely to be related to cancer. Functional roles of the top 20 candidate GBM miRNAs are described in Table 1 and Supplementary Table 19. Among them, miR-9/9* (ranked second and seventh), miR-181a/a* (ranked third and 10th), miR-21 (ranked fourth), miR-93 (ranked fifth),

Rank	miRNA	HMDD	Evidence	Reference
1	miR-22	O	Correlation between miRNA expression changes and alteration in mRNA levels of different glioma cells after PUFA or temozolomide treatment	Faragó et al.[84]
2,7	miR-9*/9	O	CAMTA1 is a novel tumor suppressor regulated by miR-9/9* in glioblastoma stem cells	Schraivogel et al.[73]
3,10	miR-181a/181a*	O	Hsa-mir-181a and hsa-mir-181b function as tumor suppressors in human glioma cells	Shi et al.[46]
4	miR-21	O	MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells	Chan et al.[75]
5	miR-93	O	MicroRNA miR-93 promotes tumor growth and angiogenesis by targeting integrin-β8	Fang et al.[77]
6	miR-34a	O	MicroRNA-34a inhibits glioblastoma growth by targeting multiple oncogenes.	Li et al.[78]
8	miR-29a	O	Over-expression of miR-29a/29b promotes apoptosis of GICs (GBM initiating cells) by inhibiting MCL1 protein expression	Aldaz, et al.[85]
9	miR-30a-5p	O	MiR-30a-5p antisense oligonucleotide suppresses glioma cell growth by targeting SEPT7	Jia et al.[49]
11	miR-125b	O	MiR-125 regulates the proliferation of glioblastoma stem cells by targeting E2F2	Wu et al.[88]
12	miR-30c	O	Effect of miR-21 and miR-30b/c on TRAIL-induced apoptosis in glioma cells	Quintavalle et al.[86]
13	miR-222	O	MiR-221 and miR-222 target PUMA to induce cell survival in glioblastoma	Zhang et al.[50]
14,18	miR-17-5p/3p	O	Stress response of glioblastoma cells mediated by miR-17-5p targeting PTEN and the passenger strand miR-17-3p targeting MDM2.	Li et al.[51]
15	miR-106a	O	MiR-106a inhibits glioma cell growth by targeting E2F1 independent of p53 status	Yang et al.[58]
16	miR-181d	O	MiR-181d acts as a tumor suppressor in glioma by targeting K-ras and Bcl-2	Wang et al.[83]
17	miR-15b	O	MiR-15b and miR-152 reduce glioma cell invasion and angiogenesis via NRP-2 and MMP-3	Zheng et al.[87]
19	let-7f	X	Let-7f Inhibits glioma cell proliferation, migration, and invasion by targeting periostin	Yan et al.[91]
20	let-7a	O	Overexpressed let-7a inhibits glioma cell malignancy by directly targeting K-ras, independently of PTEN	Wang et al.[90]

Table 1. Studies on top 20 miRNAs identified in GBM samples. Columns 1 and 2 present the ranking and name of miRNAs, respectively. In column 3, GBM-related miRNA in HMDD are marked. Columns 4 and 5 summarize representative evidence and studies supporting the role of miRNAs in GBM, respectively.

miR-34a (ranked sixth), miR-222 (ranked 13th), miR-17-5p/3p (ranked 14th and 18th), and miR-181d (ranked 16th) were reported as important GBM-related miRNAs. MiR-9/9* function as onco-miRNAs or tumor suppressors, depending on the cellular environment. Schraivogel et al.[73] reported that miR-9/9* are highly abundant in GBM stem cells and function as oncogenes by repressing tumor suppressor CAMTA1. Wu et al.[57] reported that miR-9 is upregulated in glioma patients with high WHO grade (III-IV) and represents a useful prognostic factor for overall survival. Although these studies indicated that high miR-9/9* expression promotes tumor progression, Gomez et al.[74] showed that, in GBM driven by EGFR mutation (EGFRvIII), miR-9 suppression leads to the enhanced tumor growth, because this miRNA targets FOXP1 transcription factor, and its upregulation can be oncogenic. MiR-181a/a* act as tumor suppressors, and Shi et al.[46] showed that miR-181a inhibits cell growth and invasion and induces apoptosis in glioma cells. MiR-21 was reported to be an important biomarker in GBM. Chan et al.[75] showed that miR-21 functions as an antiapoptotic factor by repressing apoptosis-related genes in GBM cells. Gabriely et al.[76] showed that miR-21 promotes the activity of matrix metalloproteinases (MMPs) by targeting MMP inhibitors, which results in the increase of tumor invasiveness. Zhou et al.[52] showed that miR-21 is involved in the regulation of EGFR and AKT pathways, and the suppression of cell growth in GBM. Furthermore, it was demonstrated[77] that miR-93 targets integrin β8, affecting integrin β8-induced cell death in GBM. MiR-34a has been reported as an important regulator in GBM, and several studies[56,78–80] reported that miR-34a functions as a tumor suppressor. MiR-34a targets several genes, such as c-Met, Notch-1, Notch-2, and CDK6, and regulates GBM-related pathways, e.g., p53 pathway. MiR-222 was shown[50] to inhibit cell apoptosis and induce cell survival through the direct targeting of p53-upregulated modulator of apoptosis (PUMA) in GBM. MiR-222 was shown to be involved in tumorigenesis[81] through the regulation of protein tyrosine phosphatase u (PTPu) expression in glioma cells. MiR-17-5p/3p were reported to regulate E2F1, PTEN, and MDM2. Srinivasan et al.[13] demonstrated that the expression of miR-17 is associated with survival time in GBM. Additionally, decreased expression level of E2F1 and cyclin D1, which represent miR-17 targets, were shown to correlate with longer patient survival. MiR-17-5p was shown to target PTEN and miR-17-3p targets MDM2 [51]. MiR-17 overexpression was shown to increase the overall survival time. MiR-181d functions as tumor suppressor in GBM, and it may represent both a predictive biomarker in temozolomide therapy and a prognostic marker regulating MGMT expression in GBM[82]. Furthermore, Wang et al.[83] showed that miR-181d functions as a tumor suppressor by regulating K-ras and Bcl-2.

Additionally, several studies[49,58,84–87] showed that miR-22 (ranked first), miR-29a (ranked eighth), miR-30a-5p (ranked ninth), miR-30c (ranked 12th), miR-106a (ranked 15th), and miR-15b (ranked 17th) are primarily associated with cell apoptosis and proliferation in GBM. MiR-125b (ranked 11th) was shown to have different functions depending on the environment, and targets E2F1 and inhibits the proliferation of CD133-positive glioma stem cells[88]. However, it was shown that miR-125b function as oncogene as well, promoting proliferation and
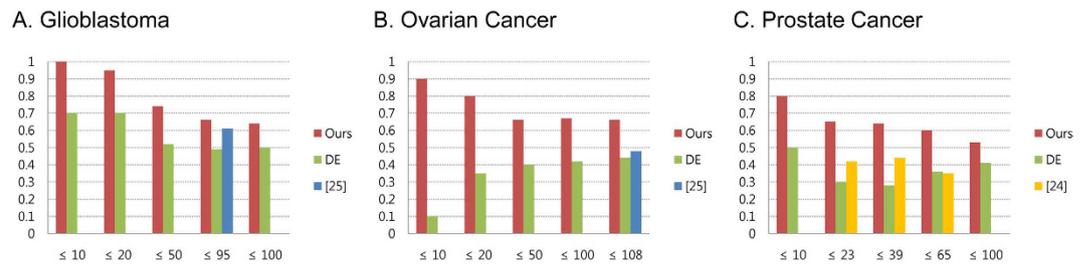
Figure 7. Comparison of our method with other methods. Our approach was compared with the approach based on DE miRNAs, Zhang et al.[24], and our previous approach[25]. Each bar represents a cumulative ratio of cancer-related miRNAs in HMDD[28] and miRNAs identified by each of the presented methods for each cancer type. The number of miRNAs identified by these methods is represented on x-axis. In (**A,B**), GBM and OVC (microarray) datasets from TCGA were used while, in (**C**), PRCA dataset (GSE34933)[24] from NCBI GEO was used for comparison.

inhibiting the apoptosis of GBM stem cells[89]. Let-7a was ranked 20th in our analysis, and Wang et al.[90] showed that it directly targets K-ras and reduces glioma cell malignancy. Although let-7f (ranked 19th) was not described as GBM-related miRNA in HMDD, Yan et al.[91] showed that it is involved in the inhibition of proliferation, migration, and invasion of glioma cells.

Functional roles of the top 20 OVC candidate miRNAs in microarray dataset are described in Supplementary Table 20. Among them, miR-21 (ranked third and 14th), miR-93 (ranked sixth), miR-20a (ranked 11th), miR-125b (ranked 13th), miR-16 (ranked 18th), and miR-27a (ranked 19th) are well-known OVC-related miRNAs. MiR-21 functions as an oncogene, and it is involved in tumorigenesis and tumor progression through the regulation of PTEN expression, an OVC-related tumor suppressor[70]. The suppression of miR-21 induces apoptosis of cisplatin-resistant OVC cells and the overexpression of miR-21 leads to the lowering of patient survival rate[67]. Similarly, miR-93 plays a key role in cell survival, since its suppression induces the apoptosis of OVC cells and it is involved in the determination of cisplatin chemosensitivity through the regulation of PTEN expression[92]. OVC cell survival was shown to be increased following miR-93 cell transfection[77]. Fan et al.[93] showed that the overexpression of miR-20a promotes proliferation and invasion through direct targeting of amyloid precursor protein (APP) in OVC cells. MiR-125b was reported to be a tumor suppressor, since it suppresses cell proliferation through the regulation of BCL3, a proto-oncogene[94], while it also inhibits tumor angiogenesis by regulating HER2 and HER3[95]. Bhattacharya et al.[96] showed that miR-16 regulates BMI-1 expression and that the downregulation of BMI-1 leads to the inhibition of proliferation and clonal growth of OVC cells. MiR-27a functions as an oncogene, and it controls the expression of multi-drug resistance (MDR)1/P-glycoprotein by targeting homeodomain interacting protein kinase (HIPK)2 that acts as a tumor suppressor[97]. Multiple studies[98–101] showed that let-7b (ranked first), miR-29a (ranked fourth), miR-30c/30e* (ranked second and 17th) and miR-30b (ranked 20th) expression levels significantly differ between OVC tissues and controls. Furthermore, let-7b and miR-30c/30e* expression changes were associated with survival time, and the copy number of the mir-30b gene was shown to be increased in OVC[102]. MiR-29b/29c (ranked fifth and 16th) are involved in the development of cisplatin resistance[103]. Li et al.[104] showed that miR-22 (ranked seventh) inhibits cell migration and invasion, and plays a key role in OVC metastasis. MiR-17 (ranked eighth in our analysis) is involved in the regulation of OVC-related pathways, suppressing LKB1-p53-p21/WAF1 pathway, which results in the induction of OVC stem cell development[69]. Yang et al.[105] showed that miR-130a (ranked 10th) is involved in drug resistance mediated by DR1/P-glycoprotein in OVC cells.

Although miR-181a-2* (ranked ninth), miR-142-3p (ranked 12th), miR-24 (ranked 15th), and miR-29c (ranked 16th) are not indicated as OVC-related miRNAs in HMDD, Parikh et al.[106] showed that miR-181a plays a crucial role in OVC progression, by promoting TGF-β-mediated epithelial-to-mesenchymal transition through the suppression of Smad7. Furthermore, the expression of miR-142-3p was shown to be highly correlated with a set of genes, including some cancer-related genes, and a set of methylation sites in OVC[107]. MiR-24 acts as a tumor suppressor, inducing apoptosis in OVC cells[108], while miR-29c was described previously[103], together with miR-29b.

Functional roles of the top 20 OVC candidate miRNAs in RNA-Seq dataset are described in Supplementary Table 21. RNA-Seq and microarray datasets contain the expression levels of pre-miRNAs and mature miRNAs, respectively, however, eight common miRNAs in top 20 miRNAs can be observed: mir-181a-2/1 (ranked first and 13th), mir-22 (ranked third), mir-93 (ranked fourth), let-7b (ranked ninth), mir-125b-1 (ranked 11th), mir-30c-2 (ranked 14th), mir-30e (ranked 18th) and mir-17 (ranked 20th).

Let-7a (ranked fifth, sixth, and eighth), mir-200c (ranked 10th), and mir-25 (ranked 17th) are known as important OVC-related miRNAs. Let-7a-2/3/1 were reported as OVC prognosis markers. The methylation of let-7a-3 was identified in epithelial OVCs, and it affects the expressions of insulin-like growth factor 2 (IGF2) and patient survival[109]. Low let-7a expression and high LIN28B expression were correlated with poorer prognosis[110]. Two studies[111,112] showed that miR-200c acts as a regulator, reducing tumor burden and increasing the sensitivity to chemotherapy by regulating TUBB3 (class III β-tubulin gene). High miR-200c expression suppresses TUBB3 expression, which ultimately prolongs survival. Zhang et al.[113] showed that the overexpression of miR-25 induces cell proliferation by directly repressing the expressions of a pro-apoptotic protein Bim in OVC. Two studies[114,115] showed that miR-92a directly represses the expression of integrin α5, and it was shown that the increased expression of this integrin leads to the poorer survival of patients with stage III OVC. Moreover, Cai et al.[116] showed that let-7e (ranked seventh) is involved in the development of cisplatin resistance in OVC, while miR-30a/99b

(ranked 12th and 15th, respectively), were shown to be differentially expressed between OVC tissues and controls[99,100]. Additionally, the functional roles of the top 20 PRCA and BRCA candidate miRNAs are described in Supplementary Tables 22 and 23.

We demonstrated here that our proposed approach outperforms the approaches based on the datasets showing differentially expressed miRNAs in the identification of cancer-related miRNAs. Therefore, our method can be widely applied to other cancer datasets and may contribute to the elucidation of cancer-related miRNA mechanisms. Furthermore, providing the ranking information, which allows the selection of a small number of highly relevant miRNAs, represent an advantage of our method.

References

1. Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **302,** 1–12 (2007).
2. Zhang, H., Li, Y. & Lai, M. The microRNA network and tumor metastasis. *Oncogene* **29,** 937–948 (2010).
3. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154–D158 (2008).
4. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37,** 495–500 (2005).
5. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
6. Xuan, P. *et al.* Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* **8,** e70204 (2013).
7. Chen, X. & Yan, G.-Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* **4,** 5501 (2014).
8. Chen, X. *et al.* WBSMDA: within and between score for miRNA-disease association prediction. Sci. *Rep.* **6,** 21106 (2016).
9. Chen, X. *et al.* HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget*, doi: 10.18632/oncotarget.11251 (2016).
10. Chen, X. *et al.* RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci. Rep.* **5,** 13877 (2015).
11. Pasquier, C. & Gardès, J. Prediction of miRNA-disease associations with a vector space model. *Sci. Rep.* **6,** 27036 (2016).
12. Iorio, M. V. *et al.* MicroRNA signatures in human ovarian cancer. *Cancer Res.* **67,** 8699–8707 (2007).
13. Srinivasan, S., Patric, I. R. P. & Somasundaram, K. A ten-microRNA expression signature predicts survival in glioblastoma. *PLoS One* **6,** e17438 (2011).
14. Zhang, J., Chong, C. C., Chen, G. G. & Lai, P. B. A seven-microRNA expression signature predicts survival in hepatocellular carcinoma. *PLoS One* **10,** e0128628 (2015).
15. Henriksen, M., Johnsen, K. B., Andersen, H. H., Pilgaard, L. & Duroux, M. MicroRNA expression signatures determine prognosis and survival in glioblastoma multiforme-a systematic overview. *Mol. Neurobi.* **50,** 896–913 (2014).
16. Gross, A. M. *et al.* Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss. *Nat. Genet.* **46,** 939–943 (2014).
17. Muniategui, A., Pey, J., Planes, F. J. & Rubio, A. Joint analysis of miRNA and mRNA expression data. *Brief. Bioinform.* **14,** 263–278 (2013).
18. McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
19. Network, C. G. A. R. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
20. Network, C. G. A. R. *et al.* The molecular taxonomy of primary prostate cancer. *Cell* **163,** 1011–1025 (2015).
21. Network, C. G. A. *et al.* Comprehensive molecular portraits of human breast tumours. N*ature* **490,** 61–70 (2012).
22. Chen, X., Slack, F. J. & Zhao, H. Joint analysis of expression profiles from multiple cancers improves the identification of microRNA–gene interactions. *Bioinformatics* **29,** 2137 (2013).
23. Zhang, S., Li, Q., Liu, J. &Zhou, X. J. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics* **27,** i401–i409 (2011).
24. Zhang, W. *et al.* Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. *J. Transl. Med.* **12,** 66 (2014).
25. Jin, D. & Lee, H. A computational approach to identifying gene-microRNA modules in cancer. PLoS Comput. *Biol.* **11,** e1004042 (2015).
26. Aerts, S. *et al.* Gene prioritization through genomic data fusion. *Nat Biotechnol.* **24,** 537–544 (2006).
27. Ma, X., Lee, H., Wang, L. & Sun, F. CGI: a new approach for prioritizing genes by combining gene expression and protein–protein interaction data. *Bioinformatics* **23,** 215–221 (2007).
28. Li, Y. *et al.* HMDD v2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **42,** D1070–D1074 (2014).
29. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302,** 249–255 (2003).
30. Hsu, S.-D. *et al.* miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **39,** D163–D169 (2010).
31. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9,** 811–818 (1990).
32. Beasley, T. M., Erickson, S. & Allison, D. B. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* **39,** 580–595 (2009).
33. Furnari, F. B. *et al.* Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes. Dev.* **21,** 2683–2710 (2007).
34. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321,** 1807–1812 (2008).
35. Kaur, M. *et al.* Database for exploration of functional context of genes implicated in ovarian cancer. *Nucleic Acids Res.* **37,** D820–D823 (2009).
36. Eisele, G. & Weller, M. Targeting apoptosis pathways in glioblastoma. *Cancer Lett.* **332,** 335–345 (2013).
37. Serão, N. V., Delfino, K. R., Southey, B. R., Beever, J. E. & Rodriguez-Zas, S. L. Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med. Genomics* **4,** 49 (2011).
38. Hao, C. *et al.* Cytokine and cytokine receptor mRNA expression in human glioblastomas: evidence of Th1, Th2 and Th3 cytokine dysregulation. *Acta Neuropathol.* **103,** 171–178 (2002).
39. Clark, P. A. *et al.* Activation of multiple ERBB family receptors mediates glioblastoma cancer stem-like cell resistance to EGFR-targeted inhibition. *Neoplasia* **14,** 420–IN13 (2012).
40. Tu, Y. *et al.* Activation of JAK/STAT signal pathway predicts poor prognosis of patients with gliomas. *Med. Oncol.* **28,** 15–23 (2011).
41. Albert, L., Karsy, M., Murali, R. & Jhanwar-Uniyal, M. Inhibition of mTOR activates the MAPK pathway in glioblastoma multiforme. *Cancer Genomics Proteomics* **6,** 255–261 (2009).
42. Nagpal, J. *et al.* Revisiting the role of p53 in primary and secondary glioblastomas. *Anticancer Res.* **26,** 4633–4639 (2006).
43. Joseph, J. V., Balasubramaniyan, V., Walenkamp, A. & Kruyt, F. A. TGF-β as a therapeutic target in high grade gliomas–promises and challenges. *Biochem. Pharmacol.* **85,** 478–485 (2013).

44. Blázquez, C. *et al.* Cannabinoids inhibit the vascular endothelial growth factor pathway in gliomas. *Cancer Res.* **64,** 5617–5623 (2004).

45. Gong, A. & Huang, S. FoxM1 and Wnt/β-catenin signaling in glioma stem cells. *Cancer Res.* **72,** 5658–5662 (2012).

46. Shi, L. *et al.* hsa-mir-181a and hsa-mir-181b function as tumor suppressors in human glioma cells. *Brain Res.* **1236,** 185–193 (2008).

47. Zhou, X. *et al.* Reduction of miR-21 induces glioma cell apoptosis via activating caspase 9 and 3. *Oncol. Rep.* **24,** 195–201 (2010).

48. Li, S.-Z. *et al.* MicroRNA-34a induces apoptosis in the human glioma cell line, A172, through enhanced ROS production and NOX2 expression. Biochem. Biophys. *Res. Commun.* **444,** 6–12 (2014).

49. Jia, Z., Wang, K., Wang, G., Zhang, A. & Pu, P. MiR-30a-5p antisense oligonucleotide suppresses glioma cell growth by targeting SEPT7. *PLoS One* **8,** e55008 (2013).

50. Zhang, C.-Z. *et al.* MiR-221 and miR-222 target PUMA to induce cell survival in glioblastoma. *Mol. Cancer* **9,** 229 (2010).

51. Li, H., Yang, B. B. *et al.* Stress response of glioblastoma cells mediated by miR-17-5p targeting PTEN and the passenger strand miR-17-3p targeting MDM2. *Oncotarget* **3,** 1653–1668 (2012).

52. Zhou, X. *et al.* Downregulation of miR-21 inhibits EGFR pathway and suppresses the growth of human glioblastoma cells independent of PTEN status. *Lab. Invest.* **90,** 144–155 (2010).

53. Yin, D. *et al.* miR-34a functions as a tumor suppressor modulating EGFR in glioblastoma multiforme. *Oncogene* **32,** 1155–1163 (2013).

54. Wang, G. *et al.* Resveratrol inhibits glioma cell growth via targeting oncogenic microRNAs and multiple signaling pathways. *Int. J. Oncol.* **46,** 1739–1747 (2015).

55. Papagiannakopoulos, T., Shapiro, A. & Kosik, K. S. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. *Cancer Res.* **68,** 8164–8172 (2008).

56. Luan, S., Sun, L. & Huang, F. MicroRNA-34a: a novel tumor suppressor in p53-mutant glioma cell line U251. *Arch. Med. Res.* **41,** 67–74 (2010).

57. Wu, Z. *et al.* Increased expression of microRNA-9 predicts an unfavorable prognosis in human glioma. *Mol. Cell. Biochem.* **384,** 263–268 (2013).

58. Yang, G. *et al.* MiR-106a inhibits glioma cell growth by targeting E2F1 independent of p53 status. *J. Mol. Med.* **89,** 1037–1050 (2011).

59. Wang, X. *et al.* PRDM1 is directly targeted by miR-30a-5p and modulates the Wnt/β-catenin pathway in a Dkk1-dependent manner during glioma growth. *Cancer Lett.* **331,** 211–219 (2013).

60. Kwak, H. *et al.* Downregulation of Spry2 by miR-21 triggers malignancy in human gliomas. *Oncogene* **30,** 2433–2442 (2011).

61. Wall, L., Burke, F., Barton, C., Smyth, J. & Balkwill, F. IFN-γ induces apoptosis in ovarian cancer cells *in vivo* and *in vitro*. *Clin. Cancer Res.* **9,** 2487–2496 (2003).

62. Ahmed, N., Riley, C., Rice, G. & Quinn, M. Role of integrin receptors for fibronectin, collagen and laminin in the regulation of ovarian carcinoma functions in response to a matrix microenvironment. *Clin. Exp. Metastasis* **22,** 391–402 (2005).

63. De Graeff, P. *et al.* The ErbB signalling pathway: protein expression and prognostic value in epithelial ovarian cancer. *Br. J. Cancer* **99,** 341–349 (2008).

64. Thomas, S., Snowden, J., Zeidler, M. & Danson, S. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br. J. Cancer.* **113,** 365–371 (2015).

65. Wen, W.-H. *et al.* p53 mutations and expression in ovarian cancers: correlation with overall survival. *Int. J. Gynecol. Pathol.* **18,** 29–41 (1999).

66. Dunfield, L. D., Dwyer, E. J. C. & Nachtigal, M. W. TGFβ-induced Smad signaling remains intact in primary human ovarian cancer cells. *Endocrinology* **143,** 1174–1181 (2002).

67. Chan, J. K. *et al.* The inhibition of miR-21 promotes apoptosis and chemosensitivity in ovarian cancer. *Gynecol. Oncol.* **132,** 739–744 (2014).

68. Li, L. *et al.* MiR-17–5p up-regulates YES1 to modulate the cell cycle progression and apoptosis in ovarian cancer cell lines. *J. Cell. Biochem.* **116,** 1050–1059 (2015).

69. Liu, T., Qin, W., Hou, L. & Huang, Y. MicroRNA-17 promotes normal ovarian cancer cells to cancer stem cells development via suppression of the LKB1-p53-p21/WAF1 pathway. *Tumour Biol.* **36,** 1881–1893 (2015).

70. Lou, Y., Yang, X., Wang, F., Cui, Z. & Huang, Y. MicroRNA-21 promotes the cell proliferation, invasion and migration abilities in ovarian epithelial carcinomas through inhibiting the expression of PTEN protein. *Int. J. Mol. Med.* **26,** 819–827 (2010).

71. Wang, E. *et al.* Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. In *Semin. Cancer Biol.*, vol. 30, 4–12 (Elsevier, 2015).

72. Gao, S. *et al.* Identification and construction of combinatory cancer hallmark–based gene signature sets to predict recurrence and chemotherapy benefit in stage II colorectal cancer. *JAMA Oncol.* **2,** 37–45 (2016).

73. Schraivogel, D. *et al.* CAMTA1 is a novel tumour suppressor regulated by miR-9/9* in glioblastoma stem cells. *EMBO J.* **30,** 4309–4322 (2011).

74. Gomez, G. G. *et al.* Suppression of microRNA-9 by mutant EGFR signaling upregulates FOXP1 to enhance glioblastoma tumorigenicity. *Cancer Res.* **74,** 1429–1439 (2014).

75. Chan, J. A., Krichevsky, A. M. & Kosik, K. S. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.* **65,** 6029–6033 (2005).

76. Gabriely, G. *et al.* MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Mol. Cell. Biol.* **28,** 5369–5380 (2008).

77. Fang, L. *et al.* MicroRNA miR-93 promotes tumor growth and angiogenesis by targeting integrin-β8. *Oncogene* **30,** 806–821 (2011).

78. Li, Y. *et al.* MicroRNA-34a inhibits glioblastoma growth by targeting multiple oncogenes. *Cancer Res.* **69,** 7569–7576 (2009).

79. Guessous, F. *et al.* microRNA-34a is tumor suppressive in brain tumors and glioma stem cells. *Cell Cycle* **9,** 1031–1036 (2010).

80. Silber, J. *et al.* miR-34a repression in proneural malignant gliomas upregulates expression of its target PDGFRA and promotes tumorigenesis. *PLoS One* **7,** e33844 (2012).

81. Quintavalle, C. *et al.* miR-221/222 overexpression in human glioblastoma increases invasiveness by targeting the protein phosphate PTPμ. *Oncogene* **31,** 858–868 (2012).

82. Zhang, W. *et al.* miR-181d: a predictive glioblastoma biomarker that downregulates MGMT expression. *Neuro Oncol.* **14,** 712–719 (2012).

83. Wang, X.-F. *et al.* MiR-181d acts as a tumor suppressor in glioma by targeting K-ras and Bcl-2. *J. Cancer Res. Clin. Oncol.* **138,** 573–584 (2012).

84. Faragó, N., Fehér, L. Z., Kitajka, K., Das, U. N. & Puskás, L. G. MicroRNA profile of polyunsaturated fatty acid treated glioma cells reveal apoptosis-specific expression changes. *Lipids. Health. Dis* **10,** 173 (2011).

85. Aldaz, B. *et al.* Involvement of miRNAs in the differentiation of human glioblastoma multiforme stem-like cells. *PLoS One* **8,** e77098 (2013).

86. Quintavalle, C. *et al.* Effect of miR-21 and miR-30b/c on TRAIL-induced apoptosis in glioma cells. *Oncogene* **32,** 4001–4008 (2013).

87. Zheng, X., Chopp, M., Lu, Y., Buller, B. & Jiang, F. MiR-15b and miR-152 reduce glioma cell invasion and angiogenesis via NRP-2 and MMP-3. *Cancer Lett.* **329,** 146–154 (2013).
88. Wu, N. *et al.* miR-125b regulates the proliferation of glioblastoma stem cells by targeting E2F2. *FEBS Lett.* **586,** 3831–3839 (2012).
89. Wu, N. *et al.* MiR-125b acts as an oncogene in glioblastoma cells and inhibits cell apoptosis through p53 and p38MAPK-independent pathways. *Br. J. Cancer* **109,** 2853–2863 (2013).
90. Wang, X.-R. *et al.* Overexpressed let-7a inhibits glioma cell malignancy by directly targeting K-ras, independently of PTEN. *Neuro Oncol.* **15,** 1491–1501 (2013).
91. Yan, S. *et al.* Let-7f inhibits glioma cell proliferation, migration, and invasion by targeting periostin. *J. Cell. Biochem.* **116,** 1680–1692 (2015).
92. Fu, X., Tian, J., Zhang, L., Chen, Y. & Hao, Q. Involvement of microRNA-93, a new regulator of PTEN/Akt signaling pathway, in regulation of chemotherapeutic drug cisplatin chemosensitivity in ovarian cancer cells. *FEBS Lett.* **586,** 1279–1286 (2012).
93. Fan, X. *et al.* miR-20a promotes proliferation and invasion by targeting APP in human ovarian cancer cells. Acta biochim. biophy. *Sin.* **42,** 318–324 (2010).
94. Guan, Y., Yao, H., Zheng, Z., Qiu, G. & Sun, K. MiR-125b targets BCL3 and suppresses ovarian cancer proliferation. *Int. J. Cancer* **128,** 2274–2283 (2011).
95. He, J. *et al.* Roles and mechanism of miR-199a and miR-125b in tumor angiogenesis. *PLoS One* **8,** e56647 (2013).
96. Bhattacharya, R. *et al.* MiR-15a and miR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res.* **69,** 9090–9095 (2009).
97. Li, Z. *et al.* MiR-27a modulates MDR1/P-glycoprotein expression by targeting HIPK2 in human ovarian cancer cells. *Gynecol. Oncol.* **119,** 125–130 (2010).
98. Nam, E. J. *et al.* MicroRNA expression profiles in serous ovarian carcinoma. *Clin. Cancer Res.* **14,** 2690–2695 (2008).
99. Resnick, K. E. *et al.* The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. *Gynecol. Oncol.* **112,** 55–59 (2009).
100. Lee, H. *et al.* MicroRNA expression in ovarian carcinoma and its correlation with clinicopathological features. *World J. Surg. Oncol.* **10,** 1 (2012).
101. Shapira, I. *et al.* Circulating biomarkers for detection of ovarian cancer and predicting cancer outcomes. *Br. J. Cancer* **110,** 976–983 (2014).
102. Zhang, L. *et al.* microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci.* **103,** 9136–9141 (2006).
103. Yu, P.-N. *et al.* Downregulation of miR-29 contributes to cisplatin resistance of ovarian cancer cells. *Int. J. Cancer* **134,** 542–551 (2014).
104. Li, J. *et al.* An inhibitory effect of miR-22 on cell migration and invasion in ovarian cancer. *Gynecol. Oncol.* **119,** 543–548 (2010).
105. Yang, L. *et al.* Altered microRNA expression in cisplatin-resistant ovarian cancer cells and upregulation of miR-130a associated with MDR1/P-glycoprotein-mediated drug resistance. *Oncol. Rep.* **28,** 592–600 (2012).
106. Parikh, A. *et al.* microRNA-181a has a critical role in ovarian cancer progression through the regulation of the epithelial–mesenchymal transition. *Nat. Commun.* **5,** 2977 (2014).
107. Andreopoulos, B. & Anastassiou, D. Integrated analysis reveals hsa-miR-142 as a representative of a lymphocyte-specific gene expression and methylation signature. *Cancer Inform.* **11,** 61 (2012).
108. Yuan, D. *et al.* Enrichment analysis identifies functional microRNA-disease associations in humans. *PLoS One* **10,** e0136285(2015).
109. Lu, L., Katsaros, D., de la Longrais, I. A. R., Sochirca, O. & Yu, H. Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. *Cancer Res.* **67,** 10117–10122 (2007).
110. Permuth-Wey, J. *et al.* LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. *Cancer Res.* **71,** 3896–3903 (2011).
111. Cittelly, D. M. *et al.* Restoration of miR-200c to ovarian cancer reduces tumor burden and increases sensitivity to paclitaxel. *Mol. Cancer Ther.* **11,** 2556–2565 (2012).
112. Prislei, S. *et al.* MiR-200c and HuR in ovarian cancer. *BMC Cancer* **13,** 72 (2013).
113. Zhang, H. *et al.* MiR-25 regulates apoptosis by targeting Bim in human ovarian cancer. *Oncol. Rep.* **27,** 594–598 (2012).
114. Ohyagi-Hara, C. *et al.* miR-92a inhibits peritoneal dissemination of ovarian cancer cells by inhibiting integrin α5 expression. *Am. J. Pathol.* **182,** 1876–1889 (2013).
115. Li, M. *et al.* miR-92a family and their target genes in tumorigenesis and metastasis. *Exp. Cell Res.* **323,** 1–6 (2014).
116. Cai, J. *et al.* Deregulation of let-7e in epithelial ovarian cancer promotes the development of resistance to cisplatin. Oncoge*nesis* **2,** e75 (2013).

Acknowledgements

Author Contributions

H.L. designed the project, D.J. and H.L. developed the algorithm and analyzed the results, and D.J. performed the experiments. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Jin, D. and Lee, H. Prioritizing cancer-related microRNAs by integrating microRNA and mRNA datasets. *Sci. Rep.* **6**, 35350; doi: 10.1038/srep35350 (2016).