Ensemble Feature Selection and Meta-Analysis of Cancer miRNA Biomarkers

Lopez-Rincon Alejandro[1*], Martinez-Archundia Marlet[2], Martinez-Ruiz Gustavo Ulises[3], Tonda Alberto[4]

1 Life Sciences and Health, CWI, Amsterdam, Netherlands
2 Laboratorio de Modelado Molecular, Bioinformática y Diseño de Fármacos. Escuela Superior de Medicina, Instituto Politecnico Nacional, Mexico City, Mexico
3 Federico Gomez Children's Hospital; School of Medicine, National Autonomous University of Mexico, Mexico City, Mexico
4 UMR 782 GMPA, Université Paris-Saclay, INRA, AgroParisTech, Paris, France

* alejandro.lopez@iscpif.fr

Abstract

The role of microRNAs (miRNAs) in cellular processes captured the attention of many researchers, since their dysregulation is shown to affect the cancer disease landscape by sustaining proliferative signaling, evading program cell death, and inhibiting growth suppressors. Thus, miRNAs have been considered important diagnostic and prognostic biomarkers for several types of tumors. Machine learning algorithms have proven to be able to exploit the information contained in thousands of miRNAs to accurately predict and classify cancer types. Nevertheless, extracting the most relevant miRNA expressions is fundamental to allow human experts to validate and make sense of the results obtained by automatic algorithms. We propose a novel feature selection approach, able to identify the most important miRNAs for tumor classification, based on consensus on feature relevance from high-accuracy classifiers of different typologies. The proposed methodology is tested on a real-world dataset featuring 8,129 patients, 29 different types of tumors, and 1,046 miRNAs per patient, taken from The Cancer Genome Atlas (TCGA) database. A new miRNA signature is suggested, containing the 100 most important oncogenic miRNAs identified by the presented approach. Such a signature is proved to be sufficient to identify all 29 types of cancer considered in the study, with results nearly identical to those obtained using all 1,046 features in the original dataset. Subsequently, a meta-analysis of the medical literature is performed to find references to the most important biomarkers extracted by the methodology. Besides known oncomarkers, 15 new miRNAs previously not ranked as important biomarkers for diagnosis and prognosis in cancer pathologies are uncovered. Such miRNAs, considered relevant by the machine learning algorithms, but still relatively unexplored by specialized literature, could provide further insights in the biology of cancer.

Author summary

MicroRNAs (miRNAs) are non-coding RNA molecules that regulate gene expression. In the last years, the under and over expression of miRNAs has been related to the diagnosis and prognosis of specific cancer types. While machine learning techniques can efficiently exploit the information contained in thousands of miRNAs to detect the

presence and typology of tumors, it is still fundamental to isolate the minimum possible number of meaningful features, in order to allow human experts to validate the results. We propose a new ensemble feature selection methodology, and we test it on a real-world dataset, taken from The Cancer Genome Atlas (TCGA) database. The considered dataset contains 1,046 miRNA expressions, data for 8,129 patients, with 29 classes of tumors. Feature selection is performed by considering the 100 most relevant features emerging from the consensus between 8 state-of-the-art classifiers with high accuracy on the dataset. Such list is shown to be sufficient to provide an unchanged classification accuracy. Finally, the 50 most important features selected by our approach are validated by human experts, resorting to a literature review. Interestingly, while most of the selected miRNAs are known oncomarkers, a few appear still understudied, and might thus represent promising leads for future research.

1 Introduction

Several studies have shown the properties of microRNA types (miRNAs) as oncogenes and tumor suppressors [1–3]. Since then, many sophisticated techniques, such as high-throughput technologies, microarray, mass spectrometry and especially the Next Generation Sequencing (NGS), have been developed for their identification [4]. However, it is clear that the development of computational tools is needed for the interpretation of results from these high-throughput experiments [5]. Indeed, computational assisted methods are used for the identification of miRNAs from different genome organisms, for example in *Caenorhabditis briggsae* [6] and in Epstein-Barr virus (EBV or HHV4), a member of the human herpesvirus (HHV) [7]. Furthermore, several computational techniques can be applied to accurately predict miRNA expressions, as seen for example in [8].

Succeeding the earliest evidence of miRNA involvement in human cancer by Croce and collaborators [9], various studies demonstrate that miRNA expression is deregulated in human cancer through diverse mechanisms [10]. Additionally, in comparison to the impractical and invasive methods currently used for cancer diagnosis [11, 12], miRNA biomarkers can be detected directly from biological fluids (such as blood, urine, saliva and pleural fluid [13]), and they can also be used as biomarkers to detect tumors at an early stage, which is extremely important for survival. For example, the 5-year survival rate for lung cancer is 5%, but an early diagnosis can boost it to almost 50% [14]. Thus, miRNA expression profiles correlate with clinical variables, highlighting their potential value as prognostic and/or diagnostic tools.

In such a context of increasing availability of data, it is of utmost practical importance to build databases of miRNA expressions data for cancer research [15–19], and also to extract features that could be used as cancer biomarkers [20–22]. For example, miRNA *hsa-mir-21* is mentioned as a marker for patients with squamous cell lung carcinoma [23], with astrocytoma [24], breast cancer [25], and gastric cancer [26]. Following this idea, the scientific community is currently looking for miRNA signatures, representing the minimal number of miRNAs to be measured for discriminating between different stages and types of cancer.

Current NGS technologies such as Applied Biosystems, SOLiD3,or HiSeq from Illumina are able to extract thousands of components in genome sequences [27], and traditional linear statistical analysis are not suited to manage such quantities of measured elements with non-lineal relationships to extract meaningful features. Thus, a suitable solution, is to use machine learning techniques for analysis, classification, and relevant features extraction of miRNA data [28–30].

Starting from a dataset containing 8,129 patients, 29 different types of cancer, and 1,046 different miRNA expressions, 8 state-of-the-art classifiers are used to extract the

most relevant miRNAs to use as biomarkers for cancer classification. Typically, classifiers trained on a dataset will not use the whole set of available features to separate classes, but just a subset which could be ordered by relative importance, with a different meaning given to the list by the specific technique. The top 100 biomarkers in the list are then evaluated as a potential reduced signature for classification. Finally, the top 50 miRNAs are compared to a meta-analysis of the medical literature, to validate the results automatically produced by the machine learning algorithms. Unsurprisingly, most of the miRNAs identified by the classifiers are also considered important by the specialized literature: 15 of them, however, are still understudied, and they could thus represent promising leads for future exploration.

The rest of the paper is organized as follows. The target dataset and the proposed approach are detailed in Section 2. Experimental results are reported in Section 3, while Section 4 concludes the paper.

2 Methods

The considered dataset, containing miRNA sequencing isoform values, is taken from the Cancer Genome Atlas[1]. The database contains the information from 8,129 patients. Using the next-generation sequencing miRNASeq BCGSC IlluminaHiSeq miRNASeq Level_3, a total of 1,046 miRNA expression features for each case study are extracted. In summary, the dataset that will be used in the following experiments has 29 types of tumors, 1,046 miRNA features, and 8,129 patient samples. Information on the dataset is summarized in Table 1.

Tumor Type	Number of Samples	Class
Adrenocortical carcinoma [ACC]	80	0
Bladder Urothelial Carcinoma [BLCA]	415	1
Breast invasive carcinoma [BRCA]	778	2
Cervical squamous cell carcinoma [CESC]	308	3
Cholangiocarcinoma [CHOL]	35	4
Esophageal carcinoma [ESCA]	47	5
FFPE Pilot Phase II [FPPP]	200	6
Head and Neck squamous cell carcinoma [HNSC]	45	7
Kidney Chromophobe [KICH]	488	8
Kidney renal clear cell carcinoma [KIRC]	66	9
Kidney renal papillary cell carcinoma [KIRP]	261	10
Lower Grade Glioma [LGG]	292	11
Liver hepatocellular carcinoma [LIHC]	530	12
Lung adenocarcinoma [LUAD]	374	13
Lung squamous cell carcinoma [LUSC]	458	14
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBC]	341	15
Mesothelioma [MESO]	87	16
Pancreatic adenocarcinoma [PAAD]	179	17
Pheochromocytoma and Paraganglioma [PCPG]	184	18
Prostate adenocarcinoma [PRAD]	500	19
Sarcoma [SARC]	262	20
Skin Cutaneous Melanoma [SKCM]	452	21
Stomach adenocarcinoma [STAD]	399	22
Testicular Germ Cell Tumors [TGCT]	155	23
Thymoma [THYM]	514	24
Thyroid carcinoma [THCA]	124	25
Uterine Carcinosarcoma [UCS]	418	26
Uterine Corpus Endometrial Carcinoma [UCEC]	57	27
Uveal Melanoma [UVM]	80	28

Table 1. Dataset: Tumor type, class label, and number of samples per class.

As a baseline comparison, a preliminary analysis of the available data is performed,

[1] http://cancergenome.nih.gov/

normalizing all the isoform expressions altogether and then quantifying the highest expressed miRNAs for each cancer tumor type. Next, the top 50 most expressed miRNAs for each tumor type are arranged in descending order. Finally, a coefficient $\sum_{i=1}^{N} coef = 1/pos_i$ where N is the number of tumor types in the dataset, that depends on miRNA's relative position, bringing the result displayed in Figure 1.

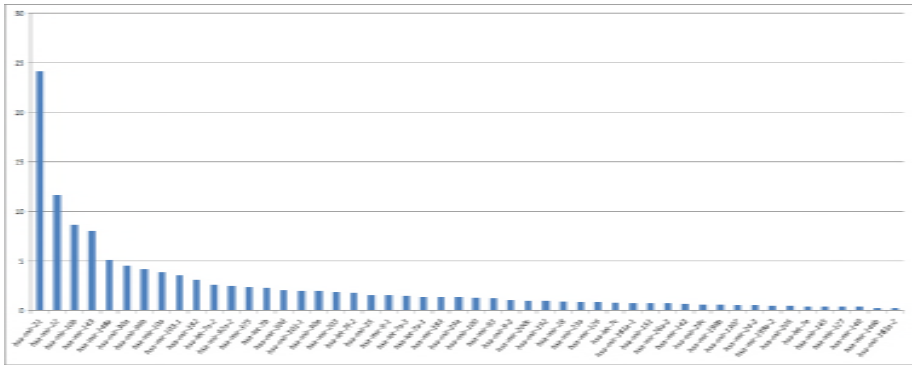

Fig 1. Top 50 most expressed miRNA types, across all cancer classes in the considered dataset.

As the objective is to find and validate a reduced list of miRNAs to be used as a signature, feature selection is to be performed on the dataset. Popular approaches to feature selection range from univariate statistical considerations, to iterated runs of the same classifier with a progressively reduced number of features, in order to assess the contribution of the features to the overall result. As the considered case study is particularly complex, however, relying upon simple statistical analyses or a single classifier might not suffice. Following the idea behind *ensemble feature selection* [31–33], we use multiple algorithms to obtain a more robust predictive performance. For this purpose, we train a set of classifiers to then extract a sorted list of the most relevant features from each. As, intuitively, a feature considered important by the majority of classifiers in the set is likely to be relevant for our aim, the information from all classifiers is then compiled to find the most common relevant features.

Starting from a thorough comparison of 22 different state-of-the-art classifiers on the considered dataset presented in [34], in this work a subset of those classifiers is selected considering both (i) high accuracy and (ii) a way to extract the relative importance of the features from the trained classifier. After preliminary tests to set algorithms' hyperparameters, 8 classifiers are chosen, all featuring an average accuracy higher than 90% on a 10-fold cross-validation:

- `BaggingClassifier` [35]
- `GradientBoostingClassifier` [36]
- `LogisticRegression` [37]
- `PassiveAggressiveClassifier` [38]
- `RandomForestClassifier` [39]
- `RidgeClassifier` [40]
- `SGDClassifier` (Stochastic Gradient Descent on linear models) [41]
- `SVC` (Support Vector Machines Classifier with a linear kernel) [42]

All considered classifiers are implemented in the `scikit-learn` Python toolbox [43].

Overall, the selected classifiers fall into two broad typologies: those exploiting ensembles of classification trees [44] (`Bagging`, `GradientBoosting`, `RandomForest`), and those optimizing the coefficients of linear models to separate classes (`LogisticRegression`, `PassiveAggressive`, `Ridge`, `SGD`, `SVC`). Depending on classifier typology, there are two different ways of extracting relative feature importance. For classifiers based on classification trees, the features used in the splits are counted and sorted by frequency, from the most to the least common. For classifiers based on linear models, the values of the coefficients associated to each feature can be used as a proxy of their relative importance, sorting coefficients from the largest to the smallest in absolute value. As the two feature extraction methods return heterogeneous numeric values, only the relative sorting of features provided by each classifier is considered. We arbitrarily decide to extract the top 100 most relevant features, so we assign to each feature f a simple score $S_f = N_t/N_c$, where N_t is the number of times that specific features appears among the top 100 of a specific classifier instance, while N_c is the total number of classifiers instances used; for instance, a feature appearing among the 100 most relevant in 73% of the classifiers used would obtain a score $S_f = 0.73$. In order to increase the generalizability of our results, each selected classifier is run 10 times, using a 10-fold stratified cross-validation, so that each fold preserves the percentage of samples of each class of the original dataset. Thus, $N_c = 80$ (8 types of classifiers, run 10 times each). The complete procedure is summarized by Algorithm 1.

Finally, the top 50 features obtained in this way are validated with a meta-analysis of the relevant literature. In a first step, reviews of miRNA types related to cancer are inspected for the presence of the extracted features. Subsequently, the PubMed database is interrogated for references containing the identified miRNA types [2], and the results are later manually analyzed by the authors.

3 Results and Discussion

Table 2 compares the classification accuracy of each classifier using the full 1,046 features, with the accuracy obtained by the same classifier using a signature composed by selected 100 features. It is interesting to notice how the accuracy is, for most cases, unchanged, providing empirical evidence that a 100-miRNA signature is enough to obtain good classification results.

Figure 2 shows a heatmap comparing the relative frequency of the overall top 100 most frequent features, for each considered classifier. As expected, not all classifiers use the same features to separate the types of cancer, and thus using their consensus proves to be more robust than just relying upon a single algorithm. It is interesting to notice that while the overall most common biomarkers appear among the top for each classifier, some classifiers make use of only a few. For example, `BaggingClassifier` and `RidgeClassifier` do not use the vast majority of the features exploited by others to discriminate between classes. A further difference between the two is that features used by `BaggingClassifier` that are also appearing in the top 100 are clearly important for the classifier, being used in almost 100% of its 10 runs; while it is noticeable how `RidgeClassifier` probably bases its discrimination on features that do not appear among the top 100. This also explains the drop in performance when `RidgeClassifier` is forced to use the top 100 features; while `BaggingClassifier` seems to be overall unaffected by the restriction (see Table 2). One classifier, `SVC`, even slightly increases its average accuracy, probably due to the fact that the search space defined by the 100-feature signature is easier to explore for its optimization procedure.

[2] Query performed on January 20th, 2018, on https://www.ncbi.nlm.nih.gov/pubmed/. The query used is `(<mir-number>[TEXT WORD]) AND ((cancer[TEXT WORD]) OR (tumor[TEXT WORD]))`.

Algorithm 1: Extracting the 100 most relevant features from the considered medical dataset.

```
Normalize dataset by feature;
Divide dataset in N folds;
Select K classifiers;
for each fold n of N do
    for each classifier k of K do
        Train classifier k_n on all folds minus n, using all features;
        Test classifier k_n on fold n;
        Obtain sorted list l_kn of features from k_n;
        Assign weight w_fnk to each f of the 1,046 features;
        for each feature f of F do
            if f is among the top 100 features in l_kn then
                w_fnk = 1
            else
                w_fnk = 0
N_c = N · K;
for each miRNA feature f do
    N_t = Σ_n^N Σ_k^K w_fnk;
    S_f = N_t / N_c;
Select 100-feature signature, from features with highest S_f;
for each fold n of N do
    for each classifier k of K do
        Train classifier k_n on all folds minus n, using signature;
        Test classifier k_n on fold n;
Compare performance of classifiers using all features and signature;
```

Table 2. Classifiers used in the experiments. For each classifier, the average accuracy and corresponding standard deviation on a 10-fold cross validation are reported, for both the complete dataset (1,046 features) and the 100 features that have been selected as the most relevant. In the case a classifier is not using standard values for its hyperparameters, the relevant variations are summarized in the corresponding column.

Classifier	Accuracy (10-fold CV)				Feature extraction method	Hyperparameters
	1,046 features		100 features			
	avg	std	avg	std		
BaggingClassifier	0.9123	0.0039	0.9104	0.0087	Decision Trees	n_estimators=300
GradientBoostingClassifier	0.9411	0.0116	0.9414	0.0100	Decision Trees	n_estimators=300
LogisticRegression	0.9308	0.0061	0.9461	0.0071	Coefficients	-
PassiveAggressiveClassifier	0.9175	0.0085	0.9107	0.0068	Coefficients	-
RandomForestClassifier	0.9324	0.0082	0.9299	0.0078	Decision Trees	n_estimators=300
RidgeClassifier	0.9064	0.0081	0.8430	0.0107	Coefficients	-
SGDClassifier	0.9160	0.0105	0.9176	0.0077	Coefficients	-
SVC	0.9314	0.0068	0.9546	0.0043	Coefficients	Linear kernel

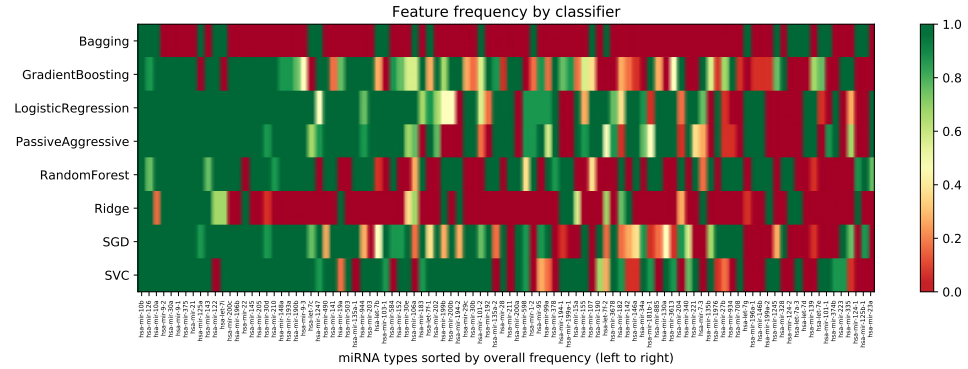

Fig 2. Heatmap with the frequency of the overall top 100 most frequent features, divided by classifier. Features are sorted by frequency, from left to right.

The top 50 features extracted, reported in Table 4, are then validated through a meta-analysis of the specialized literature. The proposed meta-analysis is mainly based on two surveys of miRNA biomarkers, Muller et al. [1], and Ferracin et al. [2], plus 150 papers obtained through a PubMed query and later manually inspected by the authors.

Of the top 50 features, 15 (*hsa-mir-10a*, *hsa-mir-21*, *hsa-mir-22*, *hsa-mir-30d*, emphhsa-mir-106a, *hsa-mir-126*, *hsa-mir-141*, *hsa-mir-143*, *hsa-mir-200b*, *hsa-mir-200c*, *hsa-mir-203*, *hsa-mir-205*, *hsa-mir-210*, *hsa-mir-375*, and *hsa-let-tc*) are mentioned either for cancer diagnosis, prediction, or prognosis by Muller [3], and can be measured by body fluids. Ferracin [2] mentions a total of 10, with 6 not appearing in [3]: *hsa-let-7b*, *hsa-let-7f-1*, *hsa-let-7i*, *hsa-mir-29c*, and *hsa-mir-145*. In total, 21 miRNAs identified by the proposed approach are found among the two surveys.

The most important result of this work is that several miRNAs types highly ranked by the classifiers appear to be still understudied in literature. Table 3 reports a summary of the literature meta-analysis, from which it clearly emerges that 15 of the miRNA types appear in less than 50 references connected to cancer from the PubMed query: strikingly, *hsa-135-a-1* appears only once, and *hsa-103-1* only twice, while both are considered important by more than 60% of the classifiers. These two miRNAs could thus represent promising leads for future research. Interestingly, only 25 of the miRNA types found by the proposed machine learning approach also appear among the most expressed in the baseline analysis summarized in Figure 1, and only 4 (*hsa-mir-9-1*, *hsa-mir-28*, *hsa-mir-103-1*, and *hsa-mir-199b*) are both overexpressed and understudied, suggesting that simple statistical approaches might not be enough to extract meaningful information from such complex data.

hsa-mir-21, mentioned in both surveys, is also the most commonly overexpressed miRNA for all classes of tumors in the statistical analysis summarized by Figure 1. *hsa-mir-10b*, the feature with the highest S_f score, while not appearing in the surveys, is currently under clinical trials as oncomarker for Glioma [45], and it is mentioned in a recent article as possible oncomarker [46]. *hsa-mir-135a-1* and *hsa-mir-135a-2*, located inside chromosomes 3 and 12, respectively, generate the same mature active sequence [47]. *hsa-mir-9-1*, *hsa-mir9-2*, and *hsa-mir9-3*, generate the oncogenic *hsa-mir-9* [48]. *hsa-mir-944* at the moment is known to decrease malignant features in gastric [49], colorectal [50] and endometrial [51] cancers.

We envision that understudied miRNAs could play a key role in the biology of cancer. On the other hand, the presence of single-nucleotide polymorphism (SNP) could affect miRNA biology due to alterations in the miRNA maturation process, their

structure, and expression levels. From these, creation and loss of miRNA targeted sites by SNPs is the most inspected area [52, 53]. However, the miRNA-mediated oncogenic transcriptional landscape could be a consequence of the SNPs presence in the seed region of mature miRNAs [54], that is involved in the molecular recognition with its targeted mRNAs. Although the presence of SNPs in seed regions seems to be negatively selected [55], in a pathological condition as cancer, it would be interesting to perform further measurements to determine whether any miRNA-related SNPs is contained in miRNA signature we propose. The complete list of the 100 extracted features is in Annex A.

Table 3. miRNA types identified by the machine learning feature extraction, appearing in less than 50 references connected to cancer in a PubMed query.

miRNA type	References related to cancer	References NOT related to cancer
hsa-mir-9-1	36	8
hsa-mir-190b	8	9
hsa-mir-9-3	20	9
hsa-mir-1247	14	8
hsa-mir-490	39	20
hsa-mir-135a-1	1	2
hsa-mir-944	16	3
hsa-mir-103-1	2	2
hsa-mir-584	23	16
hsa-mir-202	43	51
hsa-mir-199b	45	63
hsa-mir-194-2	2	3
hsa-mir-101-2	8	4
hsa-mir-135a-2	1	1
hsa-mir-28	47	65

4 Conclusions

miRNAs regulate the transcriptional landscape in a fine-tuning way. Alterations in miRNA expression profiles have serious consequences for several diseases, such as cancer. Since ectopic modulation of specific miRNAs could compromise the hallmarks of cancer, it has been proposed that cancerous miRNAs could be modulated by microRNA-based therapies. In this sense, several efforts had been achieved to generate scaffold-mediated miRNA-based delivery systems, thus exploiting the miRNA-mediated therapeutic potential. On the other hand, the altered miRNA expression profile present in cancer could be used as prognostic and/or diagnostic marker. In this respect, it has been stated that several miRNA signatures correlate with clinical outcomes, highlighting the need of a miRNA-based clinical decision-making tool. In this paper, we develop a new machine learning approach to obtain a robust, reduced miRNA signature, from a dataset containing 29 different types of cancer. A further meta-analysis of literature to validate the miRNA signature shows both well-known oncogenic and underestimated miRNA types. The results of this work could potentially be used to uncover new, promising leads of research for a better understanding of miRNA behavior. Furthermore, personal-directed anti-tumoral therapy could be achieved by measurement of a specific, minimal miRNA signature, as proposed in this work.

Table 4. Table comparing the top 50 most frequent features extracted by the machine learning algorithms with existing biomarkers references in literatures. miRNAs highlighted in gray are proposed as promising venues of research, as they only appear in less than 50 references connected to cancer in literature, according to the PubMed query performed in this study. miRNAs marked with * show considerable overexpression along all classes of tumors, as reported in Figure 1.

miRNA type	S_f	Mentioned in [3]	Mentioned in [2]	Other References
hsa-mir-10b*	1.00			[56]
hsa-mir-126*	0.96	Diagnostic		[57]
hsa-mir-10a*	0.90	Diagnostic		[58]
hsa-mir-9-2*	0.88	Diagnostic/Prognostic		[59]
hsa-mir-30a*	0.88			[60]
hsa-mir-9-1*	0.88			[61]
hsa-mir-375*	0.88			[62]
hsa-mir-21*	0.88	Diagnostic/Predictive	Oncogene	[63]
hsa-mir-125a	0.86		Tumor Suppressor	[64]
hsa-mir-143*	0.85	Diagnostic	Tumor Suppressor	[65]
hsa-mir-122	0.84			[66]
hsa-let-7i	0.84		Tumor Suppressor	[67]
hsa-mir-200c*	0.75	Diagnostic		[68]
hsa-mir-196b	0.75			[69]
hsa-mir-22*	0.75	Diagnostic		[58]
hsa-mir-145*	0.75		Tumor Suppressor	[70]
hsa-mir-205*	0.75	Diagnostic		[71]
hsa-mir-30d*	0.74	Diagnostic/Prognostic		[72]
hsa-mir-210	0.74	Diagnostic/Predictive	Oncogene	[73]
hsa-mir-148a*	0.74			[74]
hsa-mir-193a	0.74			[75]
hsa-mir-190b	0.73			[76]
hsa-mir-9-3	0.69			[61]
hsa-let-7c*	0.69	Diagnostic	Tumor Suppressor	[77]
hsa-mir-1247	0.66			[78]
hsa-mir-490	0.66			[79]
hsa-mir-141	0.65	Diagnostic/Prognostic		[80]
hsa-mir-19a	0.63			[81]
hsa-mir-503	0.63			[82]
hsa-mir-135a-1	0.63			[83]
hsa-mir-944	0.63			[84]
hsa-mir-203*	0.63	Diagnostic		[85]
hsa-let-7b*	0.61		Tumor Suppressor	[86]
hsa-mir-103-1*	0.61			[87]
hsa-mir-584	0.6			[88]
hsa-mir-152	0.59			[89]
hsa-mir-30e*	0.59			[90]
hsa-mir-106a	0.58	Diagnostic		[91]
hsa-mir-183*	0.58			[81]
hsa-let-7f-1	0.58		Tumor Suppressor	[92]
hsa-mir-202	0.56			[93]
hsa-mir-199b*	0.56			[94]
hsa-mir-200b	0.56	Diagnostic		[95]
hsa-mir-194-2	0.54			[96]
hsa-mir-29c*	0.54		Tumor Suppressor	[97]
hsa-mir-30b	0.53			[72]
hsa-mir-101-2	0.53			[98]
hsa-mir-192*	0.53			[99]
hsa-mir-135a-2	0.51			[100]
hsa-mir-28*	0.51			[101]

Acknowledgements

The results published here are based upon data generated by The Cancer Genome Atlas Research Network, and the authors would like to thank all specimen donors and research groups that contributed to this project.

Funding

This work was partially funded by INRA, France; and CONACYT, Mexico.

Annex A

List of the top 100 most relevant features identified by the proposed methodology, in order of importance: hsa-mir-10b, hsa-mir-126, hsa-mir-10a, hsa-mir-9-2, hsa-mir-30a, hsa-mir-9-1, hsa-mir-375, hsa-mir-21, hsa-mir-125a, hsa-mir-143, hsa-mir-122, hsa-let-7i, hsa-mir-200c, hsa-mir-196b, hsa-mir-22, hsa-mir-145, hsa-mir-205, hsa-mir-30d, hsa-mir-210, hsa-mir-148a, hsa-mir-193a, hsa-mir-190b, hsa-mir-9-3, hsa-let-7c, hsa-mir-1247, hsa-mir-490, hsa-mir-141, hsa-mir-19a, hsa-mir-503, hsa-mir-135a-1, hsa-mir-944, hsa-mir-203, hsa-let-7b, hsa-mir-103-1, hsa-mir-584, hsa-mir-152, hsa-mir-30e, hsa-mir-106a, hsa-mir-183, hsa-let-7f-1, hsa-mir-202, hsa-mir-199b, hsa-mir-200b, hsa-mir-194-2, hsa-mir-29c, hsa-mir-30b, hsa-mir-101-2, hsa-mir-192, hsa-mir-135a-2, hsa-mir-28, hsa-mir-211, hsa-mir-200a, hsa-mir-598, hsa-mir-1-2, hsa-mir-95, hsa-mir-99a, hsa-mir-378, hsa-mir-194-1, hsa-mir-199a-1, hsa-mir-15a, hsa-mir-155, hsa-mir-107, hsa-mir-190, hsa-let-7f-2, hsa-mir-3678, hsa-mir-182, hsa-mir-142, hsa-mir-146a, hsa-mir-34a, hsa-mir-181b-1, hsa-mir-885, hsa-mir-130a, hsa-mir-3613, hsa-mir-204, hsa-mir-340, hsa-mir-221, hsa-mir-7-3, hsa-mir-135b, hsa-mir-1976, hsa-mir-27b, hsa-mir-934, hsa-mir-708, hsa-let-7g, hsa-mir-196a-1, hsa-mir-146b, hsa-mir-199a-2, hsa-mir-1245, hsa-mir-328, hsa-mir-124-2, hsa-let-7a-3, hsa-let-7d, hsa-mir-139, hsa-let-7e, hsa-mir-101-1, hsa-mir-374b, hsa-mir-223, hsa-mir-335, hsa-mir-124-1, hsa-mir-125b-1, hsa-mir-23a.

References

1. Calore F, Lovat F, Garofalo M. Non-coding RNAs and cancer. International journal of molecular sciences. 2013;14(8):17085–17110.
2. Ferracin M, Veronese A, Negrini M. Micromarkers: miRNAs in cancer diagnosis and prognosis. Expert review of molecular diagnostics. 2010;10(3):297–308.
3. Fabbri M. Non-coding RNAs and cancer. Springer; 2013.
4. Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions. Briefings in bioinformatics. 2012;15(1):1–19.
5. Mendes N, Freitas AT, Sagot MF. Current tools for the identification of miRNA genes and their targets. Nucleic acids research. 2009;37(8):2419–2433.
6. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, et al. The microRNAs of Caenorhabditis elegans. Genes & development. 2003;17(8):991–1008.
7. Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grasser FA, et al. Identification of microRNAs of the herpesvirus family. Nature methods. 2005;2(4):269.

8. Nam JW, Shin KR, Han J, Lee Y, Kim VN, Zhang BT. Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic acids research. 2005;33(11):3570–3581.

9. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences. 2002;99(24):15524–15529.

10. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal transduction and targeted therapy. 2016;1:15004.

11. Duffy MJ. Role of tumor markers in patients with solid cancers: a critical review. European journal of internal medicine. 2007;18(3):175–184.

12. Roulston J. Limitations of tumour markers in screening. British Journal of Surgery. 1990;77(9):961–962.

13. Sauter ER, Patel N. Body fluid micro (mi) RNAs as biomarkers for human cancer. Journal of Nucleic Acids Investigation. 2011;2(1):1.

14. Society AC. Cancer facts & figures. The Society; 2008.

15. Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD. Bioinformatic tools for microRNA dissection. Nucleic acids research. 2016;44(1):24–44.

16. Bhattacharya A, Cui Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA–ceRNA interactions. Nucleic acids research. 2015;44(D1):D1005–D1010.

17. Verigos J, Magklara A. Revealing the complexity of breast cancer by next generation sequencing. Cancers. 2015;7(4):2183–2200.

18. Gomez-Rueda H, Martinez-Ledesma E, Martinez-Torteya A, Palacios-Corona R, Trevino V. Integration and comparison of different genomic data for outcome prediction in cancer. BioData mining. 2015;8(1):1.

19. Koturbash I, Tolleson WH, Guo L, Yu D, Chen S, Hong H, et al. microRNAs as pharmacogenomic biomarkers for drug efficacy and drug safety assessment. Biomarkers in medicine. 2015;9(11):1153–1176.

20. Bartels CL, Tsongalis GJ. MicroRNAs: novel biomarkers for human cancer. Clinical chemistry. 2009;55(4):623–631.

21. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids—the mix of hormones and biomarkers. Nature reviews Clinical oncology. 2011;8(8):467–477.

22. Iorio MV, Croce CM. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO molecular medicine. 2012;4(3):143–159.

23. Gao W, Shen H, Liu L, Xu J, Xu J, Shu Y. MiR-21 overexpression in human primary squamous cell lung carcinoma is associated with poor patient prognosis. Journal of cancer research and clinical oncology. 2011;137(4):557–566.

24. Zhi F, Chen X, Wang S, Xia X, Shi Y, Guan W, et al. The use of hsa-miR-21, hsa-miR-181b and hsa-miR-106a as prognostic indicators of astrocytoma. European Journal of Cancer. 2010;46(9):1640–1649.

25. Yan LX, Huang XF, Shao Q, Huang MY, Deng L, Wu QL, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. Rna. 2008;14(11):2348–2360.

26. Wang D, Fan Z, Liu F, Zuo J. Hsa-miR-21 and Hsa-miR-29 in tissue as potential diagnostic and prognostic biomarkers for gastric cancer. Cellular Physiology and Biochemistry. 2015;37(4):1454–1462.

27. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in genetics. 2008;24(3):133–141.

28. Zou Q, Mao Y, Hu L, Wu Y, Ji Z. miRClassify: an advanced web server for miRNA family classification and annotation. Computers in biology and medicine. 2014;45:157–160.

29. Wang C, Hu L, Guo M, Liu X, Zou Q. imDC: an ensemble learning method for imbalanced classification with miRNA data. Genetics and Molecular Research. 2015;14(1):123–133.

30. Yousef M, Jung S, Kossenkov AV, Showe LC, Showe MK. Naive Bayes for microRNA target predictions—machine learning for microRNA targets. Bioinformatics. 2007;23(22):2987–2992.

31. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics. 2009;26(3):392–398.

32. Saeys Y, Abeel T, Van de Peer Y. Robust feature selection using ensemble feature selection techniques. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer; 2008. p. 313–325.

33. Seijo-Pardo B, Porto-Diaz I, Bolon-Canedo V, Alonso-Betanzos A. Ensemble feature selection: Homogeneous and heterogeneous approaches. Knowledge-Based Systems. 2017;118:124–139. doi:10.1016/j.knosys.2016.11.017.

34. Rincon AL, Tonda A, Elati M, Schwander O, Piwowarski B, Gallinari P. Evolutionary Optimization of Convolutional Neural Networks for Cancer miRNA Biomarkers Classification. Applied Soft Computing. 2018;doi:https://doi.org/10.1016/j.asoc.2017.12.036.

35. Breiman L. Pasting small votes for classification in large databases and on-line. Machine Learning. 1999;36(1-2):85–103.

36. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001; p. 1189–1232.

37. Cox DR. The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B (Methodological). 1958; p. 215–242.

38. Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y. Online passive-aggressive algorithms. Journal of Machine Learning Research. 2006;7(Mar):551–585.

39. Breiman L. Random forests. Machine learning. 2001;45(1):5–32.

40. Tikhonov AN. On the stability of inverse problems. In: Dokl. Akad. Nauk SSSR. vol. 39; 1943. p. 195–198.

41. Zhang T. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In: Proceedings of the twenty-first international conference on Machine learning. ACM; 2004. p. 116.

42. Hearst MA, Dumais ST, Osman E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their Applications. 1998;13(4):18–28.

43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12(Oct):2825–2830.

44. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.

45. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. Trends in molecular medicine. 2014;20(8):460–469.

46. Zhang Y, Liao RB, Hu LL, Tong BX, Hao TF, Wu HJ. The microRNA miR-10b as a potentially promising biomarker to predict the prognosis of cancer patients: a meta-analysis. Oncotarget. 2017;8(61):104543.

47. Tribollet V, Barenton B, Kroiss A, Vincent S, Zhang L, Forcet C, et al. miR-135a inhibits the invasion of cancer cells via suppression of ERR alpha. PloS one. 2016;11(5):e0156445.

48. Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, et al. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nature cell biology. 2010;12(3):247.

49. Pan T, Chen W, Yuan X, Shen J, Qin C, Wang L. miR-944 inhibits metastasis of gastric cancer by preventing the epithelial–mesenchymal transition via MACC1/Met/AKT signaling. FEBS open bio. 2017;7(7):905–914.

50. Wen L, Li Y, Jiang Z, Zhang Y, Yang B, Han F. miR-944 inhibits cell migration and invasion by targeting MACC1 in colorectal cancer. Oncology reports. 2017;37(6):3415–3422.

51. He Z, Xu H, Meng Y, Kuang Y. miR-944 acts as a prognostic marker and promotes the tumor progression in endometrial cancer. Biomedicine & Pharmacotherapy. 2017;88:902–910.

52. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, et al. SNPs in human miRNA genes affect biogenesis and function. Rna. 2009;15(9):1640–1651.

53. Moszyska A, Gebert M, Collawn JF, Bartoszewski R. SNPs in microRNA target sites and their potential role in human disease. Open biology. 2017;7(4):170019.

54. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Human mutation. 2012;33(1):254–263.

55. Jin Y, Lee CG. Single nucleotide polymorphisms associated with microRNA regulation. Biomolecules. 2013;3(2):287–302.

56. Wang YY, Ye ZY, Zhao ZS, Li L, Wang YX, Tao HQ, et al. Clinicopathologic significance of miR-10b expression in gastric carcinoma. Human pathology. 2013;44(7):1278–1285.

57. Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. BMC cancer. 2011;11(1):374.

58. Zhang C, Wang C, Chen X, Yang C, Li K, Wang J, et al. Expression profile of microRNAs in serum: a fingerprint for esophageal squamous cell carcinoma. Clinical chemistry. 2010;56(12):1871–1879.

59. Tränkenschuh W, Puls F, Christgen M, Albat C, Heim A, Poczkaj J, et al. Frequent and distinct aberrations of DNA methylation patterns in fibrolamellar carcinoma of the liver. PloS one. 2010;5(10):e13688.

60. Tan X, Qin W, Zhang L, Hang J, Li B, Zhang C, et al. A 5-microRNA signature for lung squamous cell carcinoma diagnosis and hsa-miR-31 for prognosis. Clinical cancer research. 2011;17(21):6802–6811.

61. Bandres E, Agirre X, Bitarte N, Ramirez N, Zarate R, Roman-Gomez J, et al. Epigenetic regulation of microRNA expression in colorectal cancer. International journal of cancer. 2009;125(11):2737–2743.

62. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. International Journal of Cancer. 2010;127(12):2870–2878.

63. Asaga S, Kuo C, Nguyen T, Terpenning M, Giuliano AE, Hoon DS. Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. Clinical chemistry. 2011;57(1):84–91.

64. Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic DA, Abemayor E, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009;15(17):5473–5477.

65. Gao W, Yu Y, Cao H, Shen H, Li X, Pan S, et al. Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis. Biomedicine & Pharmacotherapy. 2010;64(6):399–408.

66. Fornari F, Gramantieri L, Giovannini C, Veronese A, Ferracin M, Sabbioni S, et al. MiR-122/cyclin G1 interaction modulates p53 activity and affects doxorubicin sensitivity of human hepatocarcinoma cells. Cancer research. 2009;69(14):5761–5767.

67. Zhang P, Ma Y, Wang F, Yang J, Liu Z, Peng J, et al. Comprehensive gene and microRNA expression profiling reveals the crucial role of hsa-let-7i and its target genes in colorectal cancer metastasis. Molecular biology reports. 2012;39(2):1471–1478.

68. Yu J, Ohuchida K, Mizumoto K, Sato N, Kayashima T, Fujita H, et al. MicroRNA, hsa-miR-200c, is an independent prognostic factor in pancreatic cancer and its upregulation inhibits pancreatic cancer invasion but increases cell proliferation. Molecular cancer. 2010;9(1):169.

69. Zhao BS, Liu SG, Wang TY, Ji YH, Qi B, Tao YP, et al. Screening of microRNA in patients with esophageal cancer at same tumor node metastasis stage with different prognoses. Asian Pacific Journal of Cancer Prevention. 2013;14(1):139–143.

70. Arndt GM, Dossey L, Cullen LM, Lai A, Druker R, Eisbacher M, et al. Characterization of global microRNA expression reveals oncogenic potential of miR-145 in metastatic colorectal cancer. BMC cancer. 2009;9(1):374.

71. Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non–small-cell lung carcinoma. Journal of clinical oncology. 2009;27(12):2030–2037.

72. Lu Y, Ryan SL, Elliott DJ, Bignell GR, Futreal PA, Ellison DW, et al. Amplification and overexpression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24. 22-q24. 23 in medulloblastoma. PloS one. 2009;4(7):e6159.

73. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. British journal of haematology. 2008;141(5):672–675.

74. Zheng G, Xiong Y, Xu W, Wang Y, Chen F, Wang Z, et al. A two-microRNA signature as a potential biomarker for early gastric cancer. Oncology letters. 2014;7(3):679–684.

75. Srinivasan S, Patric IRP, Somasundaram K. A ten-microRNA expression signature predicts survival in glioblastoma. PloS one. 2011;6(3):e17438.

76. Cazzoli R, Buttitta F, Di Nicola M, Malatesta S, Marchetti A, Rom WN, et al. microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. Journal of thoracic oncology. 2013;8(9):1156–1162.

77. Holst LMB, Kaczkowski B, Glud M, Futoma-Kazmierczak E, Hansen LF, Gniadecki R. The microRNA molecular signature of atypic and common acquired melanocytic nevi: differential expression of miR-125b and let-7c. Experimental dermatology. 2011;20(3):278–280.

78. Shi S, Lu Y, Qin Y, Li W, Cheng H, Xu Y, et al. miR-1247 is correlated with prognosis of pancreatic cancer and inhibits cell proliferation by targeting neuropilins. Current molecular medicine. 2014;14(3):316–327.

79. Gusev Y. MicroRNA profiling in cancer: A bioinformatics perspective. Pan Stanford Publishing; 2009.

80. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences. 2008;105(30):10513–10518.

81. Bandres E, Cubedo E, Agirre X, Malumbres R, Zarate R, Ramirez N, et al. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. Molecular cancer. 2006;5(1):29.

82. Wang T, Ge G, Ding Y, Zhou X, Huang Z, Zhu W, et al. MiR-503 regulates cisplatin resistance of human gastric cancer cell lines by targeting IGF1R and BCL2. Chinese medical journal. 2014;127(12):2357–2362.

83. Selcuklu S, Yakicier M, Erson A. An investigation of microRNAs mapping to breast cancer related genomic gain and loss regions. Cancer genetics and cytogenetics. 2009;189(1):15–23.

84. Nordentoft I, Birkenkamp-Demtroder K, Agerbæk M, Theodorescu D, Ostenfeld MS, Hartmann A, et al. miRNAs associated with chemo-sensitivity in cell lines and in advanced bladder cancer. BMC medical genomics. 2012;5(1):40.

85. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecologic oncology. 2008;110(1):13–21.

86. Di Fazio P, Montalbano R, Neureiter D, Alinger B, Schmidt A, Merkel AL, et al. Downregulation of HMGA2 by the pan-deacetylase inhibitor panobinostat is dependent on hsa-let-7b expression in liver cancer cell lines. Experimental cell research. 2012;318(15):1832–1843.

87. Zhao Jy, Wang F, Li Y, Zhang Xb, Yang L, Wang W, et al. Five miRNAs considered as molecular targets for predicting esophageal cancer. Medical science monitor: international medical journal of experimental and clinical research. 2015;21:3222.

88. Vriens MR, Weng J, Suh I, Huynh N, Guerrero MA, Shen WT, et al. MicroRNA expression profiling is a potential diagnostic tool for thyroid cancer. Cancer. 2012;118(13):3426–3432.

89. Hanke M, Hoefig K, Merz H, Feller AC, Kausch I, Jocham D, et al. A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. In: Urologic Oncology: Seminars and Original Investigations. vol. 28. Elsevier; 2010. p. 655–661.

90. Sugihara H, Ishimoto T, Watanabe M, Sawayama H, Iwatsuki M, Baba Y, et al. Identification of miR-30e* regulation of Bmi1 expression mediated by tumor-associated macrophages in gastrointestinal cancer. PloS one. 2013;8(11):e81839.

91. Tsujiura M, Ichikawa D, Komatsu S, Shiozaki A, Takeshita H, Kosuga T, et al. Circulating microRNAs in plasma of patients with gastric cancers. British journal of cancer. 2010;102(7):1174–1179.

92. Liang S, He L, Zhao X, Miao Y, Gu Y, Guo C, et al. MicroRNA let-7f inhibits tumor invasion and metastasis by targeting MYH9 in human gastric cancer. PLoS One. 2011;6(4):e18409.

93. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer cell. 2006;9(3):189–198.

94. Lin T, Dong W, Huang J, Pan Q, Fan X, Zhang C, et al. MicroRNA-143 as a tumor suppressor for bladder cancer. The Journal of urology. 2009;181(3):1372–1380.

95. Tang H, Kong Y, Guo J, Tang Y, Xie X, Yang L, et al. Diallyl disulfide suppresses proliferation and induces apoptosis in human gastric cancer through Wnt-1 signaling pathway by up-regulation of miR-200b and miR-22. Cancer letters. 2013;340(1):72–81.

96. Wu X, Liu T, Fang O, Leach L, Hu X, Luo Z. miR-194 suppresses metastasis of non-small cell lung cancer through regulating expression of BMP1 and p27kip1. Oncogene. 2014;33(12):1506–1514.

97. Zhao JJ, Lin J, Lwin T, Yang H, Guo J, Kong W, et al. microRNA expression profile and identification of miR-29 as a prognostic marker and pathogenetic factor by targeting CDK6 in mantle cell lymphoma. Blood. 2010;115(13):2630–2639.

98. Riquelme I, Tapia O, Leal P, Sandoval A, Varga MG, Letelier P, et al. miR-101-2, miR-125b-2 and miR-451a act as potential tumor suppressors in gastric cancer through regulation of the PI3K/AKT/mTOR pathway. Cellular oncology. 2016;39(1):23–33.

99. Silakit R, Loilome W, Yongvanit P, Chusorn P, Techasen A, Boonmars T, et al. Circulating miR-192 in liver fluke-associated cholangiocarcinoma patients: a prospective prognostic indicator. Journal of hepato-biliary-pancreatic sciences. 2014;21(12):864–872.

100. Aghanoori MR, Mirzaei B, Tavallaei M. MiRNA molecular profiles in human medical conditions: connecting lung cancer and lung development phenomena. Asian Pac J Cancer Prev. 2014;15(22):9557–9565.

101. Almeida MI, Nicoloso MS, Zeng L, Ivan C, Spizzo R, Gafa R, et al. Strand-specific miR-28-5p and miR-28-3p have distinct effects in colorectal cancer cells. Gastroenterology. 2012;142(4):886–896.