Türkiye Klinikleri Biyoistatistik Dergisi
*Turkiye Klinikleri Journal of Biostatistics* Turkiye Klinikleri J Biostat. 2021;13(1):70-81

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2020-79777

# Whole Genome microRNA Expression Data in Childhood Acute Lymphoblastic Leukemia and Evaluation of microRNA Pathways Using Fuzzy C-means

## Çocukluk Çağı Akut Lenfoblastik Lösemide Tüm Genom mikroRNA İfade Verileri ve Bulanık C-ortalamaları Kullanılarak mikroRNA Yolaklarının Değerlendirilmesi

Su ÖZGÜR[a], Mehmet Nurullah ORMAN[a], Bakiye GÖKER BAĞCA[b], Muhterem DUYU[c], Özgür ÇOĞULU[d]

[a]Department of Biostatistics and Medical Informatics, Ege University Faculty of Medicine, İzmir, TURKEY
[b]Department of Medical Biology, Ege University Faculty of Medicine, İzmir, TURKEY
[c]Department of Pediatric Health and Diseases, Medeniyet University Göztepe Training and Research Hospital, İstanbul, TURKEY
[d]Department of Medical Genetics, Ege University Faculty of Medicine, İzmir, TURKEY

*This manuscript's content is a part of the MSc. thesis of 1st author/; Su Ozgur. Number of thesis 494685.*
** This study was presented as an orally in 9th EMR and Italian Region of IBS Conference 8-12 May 2017, Thessaloniki, Greece.*
***This study was presented as an orally in 18th National and 1st International Biostatistics Congress 26-29 October 2016, Antalya, Turkey.*

**ABSTRACT Objective:** Hard clustering approaches may cause some of the relationships to be overlooked due to their nature of algorithms especially in genetic datasets. But hidden relationships can be revealed by fuzzy approaches. Purpose of this study was evaluating effect of microRNAs (miRNA) on children with acute lymphoblastic leukaemia (ALL) by using miRNA expression data obtained from bone marrow samples with sets containing different numbers of elements of fuzzy C-means (FCM). **Material and Methods:** miRNA expression levels of 43 newly diagnosed ALL patients and 14 healthy subjects were analysed via FCM. Clusters containing different numbers of miRNAs were evaluated, common properties in messenger RNA (mRNA) pathways were investigated and new pathways associated with ALL and cancer were described via miRNA target prediction tools. **Results:** Significant miRNA profile was compared to control cases. Only 46 out of 108 miRNAs were found to be significantly upregulated or downregulated. Of forty six miRNAs: 8 miRNAs were labelled as tumour suppressor (17.4%), 17 miRNAs were labelled as onco-miR (37.0%) and 21 miRNAs could not be labelled (45.6%) for hematological malignancy. Fourteen (%30.4) miRNAs were found to be apoptosis-related, 27 miRNAs were in leukemia-related (58.7%) and 15 labelled miRNAs were related with cancer pathways (32.6%). hsa-miR-181b, hsa-miR-146a, hsa-miR-155, hsa-miR-181c-5p, hsa-miR-7-1-3p, hsa-miR-708-5p onco-miRs constituted a set. These miRNAs targeted 801 common mRNAs ($p<0.05$). When this sub-cluster was searched in the literature and miRNA target prediction tools system, it was found to be involved in cancer-related pathways except ALL. **Conclusion:** Hidden relationships can be defined by fuzzy approaches and those pathways may provide guidance to open up new horizons in the field of miRNA studies.

**Keywords:** Fuzzy C means algorithm; miRNA; childhood cancers; miRNA target prediction tools

**ÖZET Amaç:** Sert kümeleme yaklaşımları, özellikle genetik veri setlerinde algoritmaların doğası gereği bazı ilişkilerin gözden kaçmasına neden olabilmektedir. Ancak gizli ilişkiler, bulanık yaklaşımlarla ortaya çıkarılabilir. Bu çalışmanın amacı, mikroRNA'ların (miRNA) akut lenfoblastik lösemili (ALL) çocuklar üzerindeki etkisini, bulanık C-ortalaması (FCM) ile elde edilen ve farklı sayıda kümeleri içeren setlerle kemik iliği örneklerinden elde edilen miRNA ekspresyon verilerini kullanarak değerlendirmektir. **Gereç ve Yöntemler:** Kırk üç yeni tanılı ALL hastası ve 14 sağlıklı çocuğun miRNA ekspresyon seviyeleri FCM ile analiz edilmiştir. Farklı sayıda miRNA içeren kümeler değerlendirilerek, mesajcı RNA (mRNA) yolaklarındaki ortak özellikler araştırılmış, ALL ve kanserle ilişkili yeni yolaklar miRNA hedef tahmin araçlarıyla tanımlanmıştır. **Bulgular:** Anlamlı miRNA düzeyleri, kontrol vakalarıyla karşılaştırılmıştır. Yüz sekiz miRNA'dan sadece 46'sının önemli ölçüde yüksek regüle veya düşük regüle edildiği bulunmuştur. Kırk altı miRNA'dan, 8 (%17,4) miRNA tümör baskılayıcı, 17 (%37,0) miRNA onco-miR olarak tanımlanırken, 21 (%45,6) miRNA hematolojik malignite için tanımlanamamıştır. Tanımlanan 14 (%30,4) miRNA apoptozla ilişkili, 27 (%58,7) miRNA lösemi ile ilişkili ve 15 (%32,6) miRNA kanser yolakları ile ilişkili saptanmıştır. hsa-miR-181b, hsa-miR-146a, hsa-miR-155, hsa-miR-181c-5p, hsa-miR-7-1-3p, hsa-miR-708-5p onko-miR'leri bir küme oluşturmaktadır. Bu miRNA'lar, 801 ortak mRNA'yı hedeflemiştir ($p<0,05$). Literatürde ve miRNA hedef tahmin araçları sisteminde, bu alt küme ilişkisi araştırıldığında, ALL dışındaki kansere bağlı yolaklarda rol oynadıkları görülmüştür. **Sonuç:** Gizli ilişkiler bulanık yaklaşımlarla tanımlanabilir ve bu yolaklar, miRNA çalışmaları alanında yeni ufuklar açmak için rehberlik sağlayabilir.

**Anahtar kelimeler:** Bulanık C ortalamalar algoritması; miRNA; çocukluk çağı kanserleri; miRNA hedef tahmin araçları

**Correspondence:** Su ÖZGÜR
Department of Biostatistics and Medical Informatics, Ege University Faculty of Medicine, İzmir, TURKEY/TÜRKİYE
**E-mail:** suozgur35@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

***Received:*** 27 Oct 2020 ***Received in revised form:*** 04 Dec 2020 ***Accepted:*** 25 Dec 2020 ***Available online:*** 29 Apr 2021

The development of microarray technology increased the capability of creating large amounts of gene expression data in recent years. Microarray studies are used to measure the expression profiles of the genes of an organism at a specific time or to genotype multiple regions of DNA by using thousands of probes complementary to the unknown sequences thus allowing parallel analysis for gene expression and gene discovery on an immobilized solid support. Large data are obtained as a result of microarray studies. It is possible to provide important information from these data sets by analysing them with advanced statistical methods such as machine learning and bioinformatics analysis.[1]

Clustering is a combination of statistical and algorithmic techniques for partitioning genes into groups, and it is often used to identify similar genes based on their expression patterns. In the classical approaches, a classical set is defined by exact boundaries. Generally, the performances of clustering algorithms decrease the possibility of overlapping among clusters on a complex structured data set. In such scenarios several observations have significant belongingness to more than one cluster, creating complexity regarding the overall cluster assignments. Therefore, classical approaches are insufficient to reveal hidden structures from the complex data sets. This can be achieved by clustering techniques from machine learning literature.

Hard clustering methods assign the sample point to a specific cluster whereas fuzzy clustering methods give a probability of assignment to all clusters. Using hard clustering techniques lead to incorrect labelling when there are no apparent clear groupings in the data set.[2]

Problems mentioned earlier for hard clustering techniques can be solved by fuzzy clustering methods. Among fuzzy based clustering methodologies, fuzzy C-means (FCM) algorithm is the most reliable and suitable choice for microarray gene expression data analysis.[3] The FCM algorithm allows that one object can belong to more than one partition (cluster) and hidden relations can be lightened by the fuzzy methods.[4]

## ROLE OF MicroRNAs IN THE GENOME

MicroRNAs (miRNA) are small, highly conserved non-coding RNA molecules involved in the regulation of gene expression.[5] miRNAs control the expression of thousands of target messenger RNA (mRNAs), with each mRNAs believed to be targeted by multiple microRNAs.[6] Many studies have shown that miRNAs have a critical effect on cellular events such as differentiation, proliferation and apoptosis in haematopoiesis. miRNAs can exhibit both oncogenic and tumour suppressive properties depending on the mRNAs which they are targeting. Decrease of expression level of tumour suppressor miRNAs lead to an increase in oncogenic expression level, thereby causing tumour formation.[7] In contrast to tumour suppressor miRNAs, oncogenic miRNAs generally act in an uncontrolled growth promoter and/or apoptotic pathway. There are also many miRNAs whose role and associations with mRNAs have not yet been identified.

Studies are continued to provide information on the disease's clinical status or response to treatment and distinguish similar tumours by comparing gene expression profiles of different tumours. To be able to practice the treatment methods related with miRNAs, their effects on cellular events need to be well described. New research findings are needed to overcome these uncertainties.[7-10] However, this type of research requires high costs. It is also important to make statistical analysis appropriate to the data in terms of uncovering hidden relationships in data and reducing costs. Therefore, fuzzy clustering approaches have become even more popular in recent years.

## FUZZY C-MEANS CLUSTERING

Different clustering algorithms have been applied to gene expression data sets such as k-means, hierarchical clustering, partitioning around medoids and *self-organizing map* algorithms. However, in hard clustering techniques, one gene belongs to only one cluster. This principle leads to some limitations for studying microarray data in spite of its plausibility in many fields of clustering analysis.[11]

Gene regulation is not a linear one-to-one process and genes can participate in different genetic networks and are frequently coordinated by a variety of regulatory mechanisms. Therefore, it is expected that

single genes can belong to more than one cluster.[12,13] The most widely applied fuzzy clustering method is the FCM algorithm. The FCM algorithm is an extension of the traditional hard k-means clustering algorithm by allowing one object belongs to more than one cluster.[14]

FCM works by assigning membership to each data point corresponding to each set center based on the distance between the cluster center and the data point. The centers of the clusters are computed based on the degree of memberships of objects. The more the data is near to the cluster center, the more its membership towards the particular cluster center and summation of membership of each data point should be equal to one.[15] For a given gene, an index closes to 1 indicates a stronger association to the cluster.

After each iteration membership and cluster, the centers are updated according to the below formula: Partition matrix, U = [uij], i = 1, 2, . . . c and j = 1, 2, . . . , N are constructed

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

where, N is the number of points, m is real number -greater than 1, $u_{ij}^{m}$ is the degree of freedom of membership of $x_i$ in the cluster, j, $x_i$ is the $i^{th}$ of d-dimensional measured data, $c_j$ is the d-dimensional center of the cluster and ||...|| is the norm of similarity between any measured data center. This algorithm is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \|x_i - c_j\|^2$$

where, $\|x_i - c_j\|$ is the Euclidean distance between $i^{th}$ data and $j^{th}$ cluster center ($2 \leq c < N$). Iteration will stop when we obtain, $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon$

where $\varepsilon$ is a termination criterion between 0 and 1, whereas $k$ are the iteration steps. This procedure converges to a local minimum or a saddle point of $J_m$ (Figure 1).

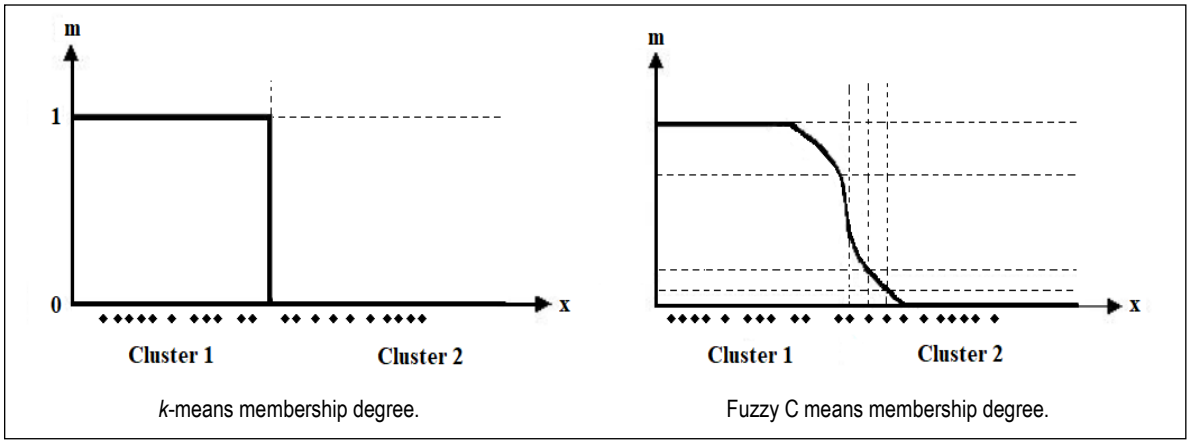


**FIGURE 1:** Membership degrees for classic and fuzzy clustering.

## OBJECTIVES

In this study, miRNA expression levels of a study conducted by Duyu et al. and supported by TUBITAK were used. They aimed to identify the relevant miRNAs in the diagnosis of childhood acute lymphoblastic leukemia (ALL) and to evaluate the effects on disease pathogenesis.[16]

The purpose of this study was to evaluate the effect of microRNAs on children with ALL by using microRNA expression data obtained from bone marrow samples with sets containing different numbers of elements of FCM algorithms and to predict genes involved in the same pathways.

## MATERIAL AND METHODS

### DATASET

In this study, we have analyzed miRNA expression levels of 43 newly diagnosed ALL patients and 14 healthy control subjects. The ages of the patients were between newborns and 18-year olds. They were selected from the newly diagnosed ALL patients in Ege University, Department of Pediatric Hematology and Department of Pediatric Oncology, Dokuz Eylul University, Department of Hematology, Dr. Behcet Uz Children's Oncology Hospital and Hematology Clinic. Individuals in the control group were selected from the same centers with normal bone marrow aspiration results.[16]

### ETHICAL APPROVAL AND DECLARATION OF HELSINKI PRINCIPLES

Board of Ethical Committee of Ege University approved the study design before starting the study (decision no: 16-6/18). This research was conducted in accordance with the Declaration of Helsinki principles.

### STATISTICAL ANALYSIS

This study has focused on evaluating the effect of miRNAs on children with ALL by using miRNA expression data obtained from bone marrow samples with sets containing different numbers of elements of FCM algorithms. This data set contains expression levels of 1,078 miRNAs of 43 ALL and 14 control cases.[16] Microarray data sets were formatted in a high dimensional matrix. Gene selection was done before fuzzy clustering because of the noisy data structure. Genes associated with ALL were selected according to the gene expression level changes (such as using 2-fold expression change as cut off). We tried to eliminate the effect of imbalance between the ALL and healthy groups by selecting genes that are up or down-regulated due to a significant 2-fold expression change. Therefore, the effect of imbalanced class distribution was decreased on the clustering analysis. Student's *t-test* was used to compare the differences in serum miRNA concentrations between the study (ALL) and control groups. Only 46 out of 108 miRNAs were found to be significantly upregulated or downregulated. Statistical significance was defined as $p<0.05$ for all comparisons. FCM clustering was performed -**e1071** package- for 46 miRNAs by using R Project for Statistical Computing (http://www.R-project.org/). Number of fuzzy clusters was named as 2, 3, 4, 5 and 6 clusters which were obtained via FCM algorithm. Each sub-group was evaluated to identify their characteristics and relationships with each other by using miRNA prediction tools. Also, infrequent miRNAs in subsets were not excluded from the analysis due to their clinical significance. miRNAs in each cluster were loaded to databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG), Online Mendelian Inheritance in Man (OMIM), miRWalk, and TargetScan to find a common effect on mRNAs. The number of mRNAs in which miRNAs have common effects is presented as support vector regression (mirSVR) score <0.05 ($p<0.05$ indicates significance level), as indicated in the system query.[17]

### TARGET PREDICTION TOOLS

#### miRWalk

miRWalk is an open-source platform providing predicted data obtained with a machine learning algorithm including experimentally verified miRNA-target interactions.[18]

### Kyoto Encyclopedia of Genes and Genomes

KEGG is an integrated database resource consisting of eighteen databases for understanding high-level functions and utilities of the biological system: Cell, the organism, and the ecosystem, from genomic and molecular-level information.[19]

### Online Mendelian Inheritance in Man

OMIM is an exhaustive catalogue of human genes and genetic disorders and traits, with a particular focus on the gene-phenotype relations that is updated daily and free.[20]

### TargetScan

TargetScan predicts biological targets of miRNAs by searching for the presence of sites that match the seed region of each miRNA and provides accurate rankings of the predicted targets for each miRNA.[21]

## RESULTS

According to the results of FCM for 46 miRNAs: 8 miRNAs were labelled as tumour suppressor (17.4%), 17 miRNAs were labelled as onco-miR (37.0%) and 21 miRNAs could not be labelled (45.6%) for hematological malignancy. Fourteen (30.4%) miRNAs were in apoptosis-related pathways, 27 miRNAs were in leukemia-related pathways (58.7%) and 15 labelled miRNAs were related with cancer pathways (32.6%) (Table 1).

When 46 miRNAs were splitted into two clusters via FCM algorithm, hsa-miR-181a, hsa-miR-195, hsa-miR-146a and hsa-miR-128 onco-miRs were grouped in the first cluster. MiRNAs constituting the first cluster were reported to target 1,613 transcripts ($p<0.05$). However, there was no evidence that this structure existed in common transcripts and associated pathways identified in ALL.[17-19] The second cluster contained 13 onco-miRs, 8 tumor suppressor-miRs and 21 unidentified miRNAs. This set did not exist in known ALL pathways as well.

When the subsets obtained as a result of triple clustering were evaluated; hsa-miR-181a-5p, hsa-miR-195-5p and hsa-miR-128-3p onco-miRs were included in the first cluster. These three miRNAs seemed to target 2795 mRNAs ($p<0.05$).[17] Among these miRNAs, ALL-related transcripts could not be identified.[18,20] Similarly, no information was found that this structure was involved in common transcripts and pathways associated with ALL.[18,19] Hsa-miR-181b, hsa-miR-146a, hsa-miR-155, hsa-miR-181c-5p, hsa-miR-7-1-3p, hsa-miR-708-5p onco-miRs constituted the second cluster. These miRNAs targeted 801 common mRNAs ($p<0.05$).[17] When this sub-cluster was searched in the literature and miRNA target prediction tools system, it was found to be involved in cancer-related pathways except ALL.[18,20] The third cluster contained 8 onco-miRs and 8 tumor suppressor-miRs and 21 unidentified miRNAs. Although those miRNAs targeted common transcripts, there was no information about their association with ALL pathways.

When the FCM clustering algorithm was applied and the structure was splitted into 4 subgroups, hsa-miR-181a was clustered in the first cluster and hsa-miR-195 was in the second sub-cluster. miRNAs were searched by miRNA-target prediction tools; hsa-miR-181a was found to be involved in cancer-related pathways and hsa-miR-195 was involved in leukemia-related pathways. Hsa-miR-181b, hsa-miR-146a, hsa-miR-155, hsa-miR-1322, hsa-miR-128, hsa-miR-181c, hsa-miR-7-1* and hsa-miR-708 constituted the third cluster and they were targeted to 461 common mRNAs. There was no ALL-associated transcript targeted by all those 8 miRNAs.[18,20] In the fourth cluster, there were 36 miRNAs, in which eight were identified as onco-miRs and eight were tumour suppressor-miRNAs. No transcript was identified which allowed them to be distinguished in common.

**TABLE 1:** Characteristics of microRNAs.

| miRNA | Characteristics | Apoptosis | Leukemia | Pathways |
|---|---|---|---|---|
| hsa-miR-181a | OncomiR | x | x | Cancer pathways |
| hsa-miR-146a | OncomiR | x | x | x |
| hsa-miR-195 | OncomiR | x | Leukemia | x |
| hsa-miR-181b | OncomiR | x | x | Cancer pathways |
| hsa-miR-155 | OncomiR | Apoptosis | Leukemia | x |
| hsa-miR-181a* | OncomiR | x | x | Cancer pathways |
| hsa-miR-181c | OncomiR | x | Leukemia | x |
| hsa-miR-7-1* | OncomiR | x | Leukemia | x |
| hsa-miR-708 | OncomiR | x | x | x |
| hsa-miR-1,322 | x | Apoptosis | Leukemia | x |
| hsa-miR-297 | x | x | Leukemia | x |
| hsa-miR-185 | x | x | Leukemia | x |
| hsa-miR-128 | OncomiR | Apoptosis | Leukemia | x |
| hsa-miR-145 | Tumour supressor-miRNA | x | Leukemia | x |
| hsa-miR-1,323 | OncomiR | x | Leukemia | x |
| hsa-let-7b | OncomiR | x | Leukemia | x |
| hsa-miR-4,312 | x | x | Leukemia | x |
| hsa-miR-369-3p | x | x | x | x |
| hsa-miR-181a-2* | x | x | x | Cancer pathways |
| hsa-miR-548i | OncomiR | x | Leukemia | Cancer pathways |
| hsa-miR-3,173 | OncomiR | Apoptosis | Leukemia | Cancer pathways |
| hsa-miR-640 | Tumoursupressor-miRNA | Apoptosis | x | Cancer pathways |
| hsa-miR-181d | OncomiR | Apoptosis | Leukemia | Cancer pathways |
| hsa-miR-769-3p | x | Apoptosis | Leukemia | x |
| hsa-miR-147b | Tumoursupressor-miRNA | x | x | x |
| hsa-miR-1,284 | Tumoursupressor-miRNA | x | x | x |
| hsa-miR-580 | Tumoursupressor-miRNA | Apoptosis | x | x |
| hsa-miR-3,174 | x | Apoptosis | Leukemia | x |
| hsa-miR-924 | x | Apoptosis | Leukemia | Cancer pathways |
| hsa-miR-199b-5p | tumoursupressor-miRNA | x | Leukemia | x |
| hsa-miR-606 | tumoursupressor-miRNA | x | x | Cancer pathways |
| hsa-miR-2,115 | x | x | Leukemia | x |
| hsa-miR-27a | tumoursupressor-miRNA | x | Leukemia | x |
| hsa-miR-29b-1* | x | Apoptosis | x | Cancer pathways |
| hsa-miR-3,115 | x | x | x | x |
| hsa-miR-3,121 | OncomiR | x | x | x |
| hsa-miR-3,140 | OncomiR | Apoptosis | x | Cancer pathways |
| hsa-miR-337-3p | x | x | x | Cancer pathways |
| hsa-miR-4,320 | x | x | Leukemia | x |
| hsa-miR-548w | x | x | x | x |
| hsa-miR-105 | x | Apoptosis | Leukemia | x |
| hsa-miR-1,178 | x | x | Leukemia | Cancer pathways |
| hsa-miR-92a-1* | x | x | Leukemia | x |
| hsa-miR-1,321 | x | x | Leukemia | Cancer pathways |
| hsa-miR-1,179 | x | x | x | x |
| hsa-miR-1,265 | x | Apoptosis | Leukemia | x |

miRNA: MicroRNA.

When all the samples were splitted into five clusters; hsa-miR-181a alone constituted the first cluster. hsa-miR-195 and hsa-miR-128 were found to co-target the lymphoblastic leukemia 1 (LYL1) transcript validated for ALL.[18,19] Moreover, this structure, which had a common role in leukocyte receptor signalling and leukemia pathways, constituted the second cluster.[18,20] In the cluster of hsa-miR-181b, hsa-miR-146a, hsa-miR-155, hsa-miR-181c and hsa-miR-7-1*, miRNAs that exhibited oncomiR properties were aggregated together. This structure appeared to be involved in pathways associated with leukemia. hsa-miR-181a*, hsa-miR-1322, hsa-miR-297, hsa-miR-708 constituted the fourth cluster. These miRNAs had a common transcript validated in cancer pathways. In the fifth cluster, 34 miRNAs were existing and none of them were involved in any known pathway associated with ALL.

In the last clustering phase, the structure was splitted into 6 sub-clusters. The first sub-cluster contained 2 miRNAs which were oncomiRs and had 3,545 common targets (mRNAs). hsa-miR-181a and hsa-miR-146a miRNAs in the cluster were associated with cancer pathways (exp. gastric cancer, papillary thyroid carcinoma). On the other hand, this set was not existing in the known ALL pathways. The second cluster contained miR-195-5p which had been identified as oncomiR and effected 8,843 mRNAs. This miRNA was associated with ALL pathways.

The third set contained 6 miRNAs which acted as oncomiR and they had effects on 219 mRNAs. When they were searched in miRNA target prediction tools (KEGG, OMIM, miRWalk, TargetScan), it was detected that miRNAs were involved in cancer-related pathways. However, this set was not existing in known ALL pathways. The fourth cluster included 3 miRNAs which had unknown characteristics and effected 1,498 mRNAs which were involved in cancer-related pathways.

The fifth cluster contained only one miRNA which had been characterized as oncomiR. It effected 8,711 mRNAs and was related with ALL pathways. The sixth cluster had 33 miRNAs and their effects were observed on 5 mRNAs. Eight of them were specified as tumour suppressor, 7 of them were oncomiR and the others were unspecified. When searched in miRNA target prediction tools (KEGG, OMIM, miRWalk, TargetScan), those miRNAs were not existing in known ALL pathways (Table 2), (Figure 2).

Means of expression levels of clusters were presented in Table 3. The minimum and maximum number of clusters were determined according to the clinical experience of the physicians. Since some clusters contained single miRNAs, the inter-cluster means were not compared.

## DISCUSSION

One of the first discovered oncogenic miRNAs is miR-155 and target mRNA of miR-155 has not been fully identified. Expression level of miR-155 has been found to be high in Hodgkin's lymphoma, lung, breast, and pancreatic cancers.[22-25] The interaction of hsa-miR 146a-5p targeting 5448 mRNAs, especially the NUP214 gene, with the Coding Domain Sequence (CDS) region is closely associated with ALL.[18,20] hsa-miR-181a is a member of the miR-181 family and regulates 7024 mRNAs.[20,21] miR-181 family plays a role in leukocyte differentiation, maturation, and function.[26] However, miR-181a-5p has been found to provide tumour-suppressing function in various types of cancers by providing multiple critical gene regulations.[27-30] In relation to ALL, it has been shown that miR-181a-5p may interact with *CDS*, Rhombotin (LMO)-1 and pre-B-cell leukemia homeobox 1 (*PBX1)* 5 5 untranslated (UTR) region regions of the LMO2 gene.[18,20] hsa-miR-195 is a member of the mir-15 precursor family.[21] There are many studies indicating that this family is associated with the development of leukemia, especially chronic lymphoblastic leukemia (CLL).[31-33] Although the expression level of hsa-miR-195-5p, which is a tumour -suppressive miRNA, has been shown to decrease in different types of cancers, it has been reported that its expression level increases in ALL and CLL.[34-38] hsa-miR-195-5p regulates 6,985 mRNAs and among them, especially the regulation of the 5'UTR region of the *LYL1* gene plays a role in the pathogenesis of ALL.[17-20] hsa-miR-181b is a member of the miR-181 family and targets 7,187 mRNAs.[17] The interaction of LMO2 gene with *CDS*, *LMO1* and *PBX1*

**TABLE 2:** Evaluation of miRNAs via fuzzy C means algorithm and common targeted mRNAs.

| miRNA | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|
| hsa-miR-181a | 1 | 1 | 1 | 1 | 1 |
| hsa-miR-195 | 1 | 1 | 2 | 2 | 2 |
| hsa-miR-181b | 2 | 2 | 3 | 3 | 3 |
| hsa-miR-146a | 1 | 2 | 3 | 3 | 1 |
| hsa-miR-155 | 2 | 2 | 3 | 3 | 3 |
| hsa-miR-145 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-181a* | 2 | 3 | 4 | 4 | 3 |
| hsa-miR-1,322 | 2 | 3 | 3 | 4 | 4 |
| hsa-miR-1,323 | 2 | 3 | 4 | 5 | 6 |
| hsa-let-7b | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-4,312 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-369-3p | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-181a-2* | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-548i | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-128 | 1 | 1 | 3 | 2 | 5 |
| hsa-miR-3,173 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-640 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-181d | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-181c | 2 | 2 | 3 | 3 | 3 |
| hsa-miR-7-1* | 2 | 2 | 3 | 3 | 3 |
| hsa-miR-769-3p | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-147b | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-1,284 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-580 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-3,174 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-924 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-199b-5p | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-606 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-297 | 2 | 3 | 4 | 4 | 4 |
| hsa-miR-708 | 2 | 2 | 3 | 4 | 3 |
| hsa-miR-2,115 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-27a | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-29b-1* | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-3,115 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-3,121 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-3,140 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-337-3p | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-4,320 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-548w | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-105 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-1,178 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-92a-1* | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-1,321 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-185 | 2 | 3 | 4 | 5 | 4 |
| hsa-miR-1,179 | 2 | 3 | 4 | 5 | 6 |
| hsa-miR-1,265 | 2 | 3 | 4 | 5 | 6 |

| Cluster 6 | Number of miRNA | Count of targeted common mRNA | mirSVR score p value |
|---|---|---|---|
| Group 1 | 2 | 3,545 | <0.001 |
| Group 2 | 1 | 8,843 | <0.001 |
| Group 3 | 3 | 1,498 | <0.001 |
| Group 4 | 6 | 219 | <0.001 |
| Group 5 | 1 | 8,711 | <0.001 |
| Group 6 | 33 | 5 | <0.001 |

| Cluster 5 | Number of miRNA | Count of targeted common mRNA | mirSVR score p value |
|---|---|---|---|
| Group 1 | 1 | 7,847 | <0.001 |
| Group 2 | 2 | 4,912 | <0.001 |
| Group 3 | 5 | 801 | <0.001 |
| Group 4 | 4 | 232 | <0.001 |
| Group 5 | 34 | 5 | <0.001 |

| Cluster 4 | Number of miRNA | Count of targeted common mRNA | mirSVR score p value |
|---|---|---|---|
| Group 1 | 1 | 8,843 | <0.001 |
| Group 2 | 1 | 7,847 | <0.001 |
| Group 3 | 8 | 461 | <0.001 |
| Group 4 | 36 | 2 | <0.001 |

| Cluster 3 | Number of miRNA | Count of targeted common mRNA | mirSVR score p value |
|---|---|---|---|
| Group 1 | 3 | 2,795 | <0.001 |
| Group 2 | 6 | 801 | <0.001 |
| Group 3 | 37 | 1 | <0.001 |

| Cluster 2 | Number of miRNA | Count of targeted common mRNA | mirSVR score p value |
|---|---|---|---|
| Group 1 | 4 | 1,613 | <0.001 |
| Group 2 | 42 | ** | ** |

miRNA: MicroRNA; mRNA; Messenger RNA; p<0.05 Significance level; mirSVR: micro RNA support vector regression.

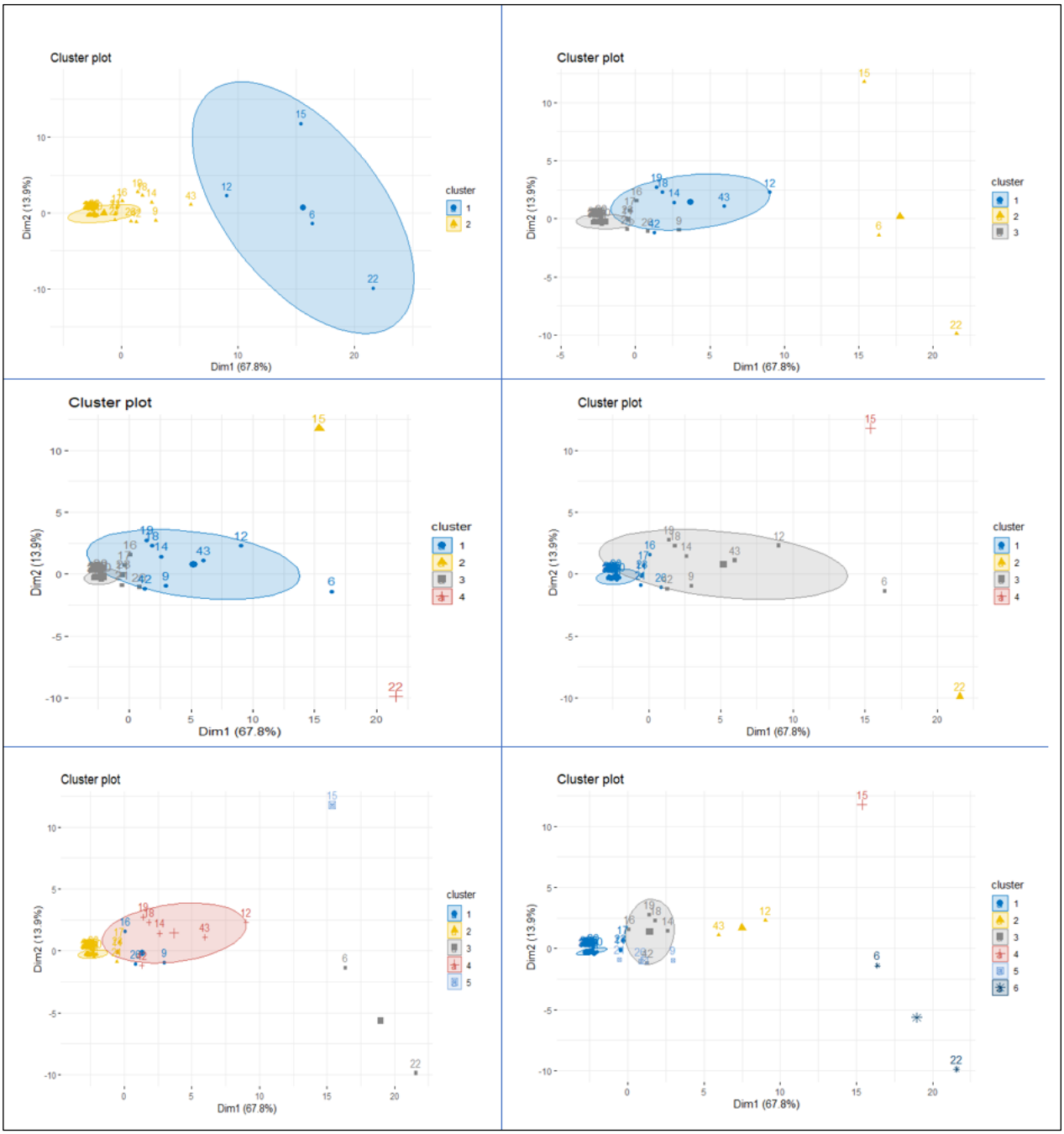


**FIGURE 2.** Evaluation of microRNAs via fuzzy C means algorithm.

**TABLE 3:** Means of expression levels by clusters.

| Number of cluster | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| 2 | 1,727.83 | 132.1179 | | | | |
| 3 | 1,925.197 | 659.2461 | 89.16972 | | | |
| 4 | 1,810.676 | 2,251.429 | 728.0981 | 89.22914 | | |
| 5 | 1,810.676 | 1,982.457 | 659.2461 | 384.0192 | 104.6585 | |
| 6 | 1,473.202 | 2,251.429 | 491.2408 | 310.9066 | 1,713.485 | 50.5693 |

genes with 5 'UTR regions (same as miR-181a-5p) has been reported to be associated with ALL.[18,20] Although upregulation of hsa-miR-181b in solid tumours is involved in the induction of apoptosis, it has been reported that it may show oncogenic properties for ALL.[39,40]

The effects of miRNAs with different mechanisms in similar malignancies on ALL have not been fully elucidated. All the miRNAs which have different mechanisms and similar malignancies have not been identified for acute lymphoblastic leukemia (ALL) yet. Hard clustering approaches will cause erroneous evaluations due to unidentified miRNAs. In this study, miRNAs were clustered according to the expression levels by FCM algorithm which allowed us to obtain clusters involving common miRNAs. Some of these clusters were located in known pathways associated with ALL, but some of them could not be identified. The investigation of miRNAs obtained from the fuzzy approach and found in common subsets in new genetic studies is important to reveal new disease relationships. For instance, hsa-miR-1322 is a miRNA belongs to the miR-1,322 family, which has an oncogenic feature in oesophageal cancer, and it targets 5,265 mRNAs. hsa-miR-297 belongs to the miR-297 family and targets 4,662 mRNAs.[21] Nevertheless, there is no reference to show the relationship of this miRNA with leukemia**.** hsa-miR-185 belongs to mir-185 family and targets 6,938 mRNAs.[17,21] 3'UTR regions of nibrin and Transcription Factor 3 genes; coding sequence (CDS) of the LYL1 gene, potential mating of the TAL1 gene to the 5'UTR region has been associated with ALL.[18,20] However, there are no studies showing the association of this miRNA with ALL in the literature as well.

As the number of clusters increases in genetic data, it is difficult to capture the biological similarities. Because the relationships between miRNAs are not linear. Similar miRNA groups are assigned to different clusters as the number of clusters increases (due to small expression differences). Therefore, the number of clusters in our study was limited to a maximum of 6.[41]

Mean expression level of miRNAs by clusters have been reported on Table 3. These numbers indicate the performance of FCM on partitioning process. When evaluating different clusters obtained through the algorithm, regardless of selected sample and cluster size, we have found that tumour suppressor miRNAs were always together, particularly, in the most populated clusters. We can deduct from this study that this method has the potential to collect all the tumour suppressors in the same cluster.

## CONCLUSION

The use of clustering techniques has increased due to the difficulties encountered in the complexity of biological networks, the volume of genes present and challenges of comprehending and interpretation the resulting the mass of data. Classical clustering approaches may cause some of the relationships to be overlooked due to their nature of algorithms. Hard clustering approaches, especially in genetic data, will lead researchers to make incorrect evaluations. Therefore, hidden relationships can be revealed by fuzzy approaches.

In conclusion, according to our study's results, new correlations between miRNAs and mRNAs associated with ALL via the results from the data obtained from the use of this algorithm can be revealed. Potential new pathways may provide guidance to open up new horizons for miRNA studies.

***Source of Finance***

*The authors would like to thank the Scientific and Technological Research Council of Turkey (TUBITAK) for financial support under Grant 109S076.*

***Conflict of Interest***

*No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.*

### *Data Availability*

*The data that support the findings of this study are available on request from the corresponding author, (Su Özgür). The data are not publicly available due to ethical committee decision.*

### *Authorship Contributions*

***Idea/Concept:*** *Su Özgür, Mehmet Nurullah Orman, Özgür Çoğulu;* ***Design:*** *Su Özgür, Mehmet Nurullah Orman;* ***Control/Supervision:*** *Mehmet Nurullah Orman, Özgür Çoğulu;* ***Data Collection and/or Processing:*** *Muhterem Duyu, Özgür Çoğulu;* ***Analysis and/or Interpretation:*** *Su Özgür, Mehmet Nurullah Orman, Bakiye Göker Bağca, Muhterem Duyu, Özgür Çoğulu;* ***Literature Review:*** *Su Özgür, Bakiye Göker Bağca, Muhterem Duyu, Mehmet Nurullah Orman, Özgür Çoğulu;* ***Writing the Article:*** *Su Özgür, Bakiye Göker Bağca, Muhterem Duyu, Özgür Çoğulu, Mehmet Nurullah Orman;* ***Critical Review:*** *Özgür Çoğulu, Mehmet Nurullah Orman;* ***References and Fundings:*** *Özgür Çoğulu;****Materials:*** *Muhterem Duyu, Özgür Çoğulu.*

## REFERENCES


1. Greene CS, Tan J, Ung M, Moore JH, Cheng C. Big data bioinformatics. Journal of Cellular Physiology. 2014;229(12):1896-1900. [Crossref] [PubMed] [PMC]
2. Vardhan A, Sarmah P, Das A. A Comprehensive Analysis of the Most Common Hard Clustering Algorithms. In: Smys S, Bestak R, Rocha Á, eds. Inventive Computation Technologies. ICICIT 2019. Lecture Notes in Networks and Systems, Springer, Cham. vol 98. 2020. [Crossref]
3. Scaria T, Stephen G, Mathew J. Gene expression data analysis using fuzzy C-means clustering technique. International Journal of Computer Applications. 2016;135(8):33-6. [Crossref]
4. Li X, Lu X, Tian J, Gao P, Kong H, Xu G, et al. Application of fuzzy c-means clustering in data analysis of metabolomics. Anal Chem. 2009;1;81(11):4468-75. [Crossref] [PubMed]
5. Shenouda SK, Alahari SK. MicroRNA function in cancer: oncogene or a tumor suppressor? Cancer Metastasis Rev. 2009;28(3-4):369-78. [Crossref] [PubMed]
6. Pillai RS. MicroRNA function: multiple mechanisms for a tiny RNA? RNA. 2005;11(12):1753-61. [Crossref] [PubMed] [PMC]
7. Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB, et al. Common features of microRNA target prediction tools. Front Genet. 2014;18;5:23. [Crossref] [PubMed] [PMC]
8. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie/Chemical Monthly. 1994;125(2):167-88. [Crossref]
9. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39(Database issue):D152-7. [Crossref] [PubMed] [PMC]
10. Laganà A, Forte S, Giudice A, Arena MR, Puglisi PL, Giugno R, et al. miRò: a miRNA knowledge base. Database (Oxford). 2009;2009:bap008. [Crossref] [PubMed] [PMC]
11. Wang YF, Yu ZG, Anh V. Fuzzy C-means method with empirical mode decomposition for clustering microarray data. Int J Data Min Bioinform. 2013;7(2):103-17. [Crossref] [PubMed]
12. Macneil LT, Walhout AJM. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Research. 2011;21:645-57. [Crossref] [PubMed] [PMC]
13. Lloyd A. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins (Methods of Biochemical Analysis, 43). Briefings in Bioinformatics. 2001;2(4):407-8. [Crossref]
14. Bezdek J. Corrections for "FCM: the fuzzy C-means clustering algorithm." Computers & Geosciences. 1985;11(5):660. [Crossref]
15. Kim SY, Lee JW, Bae JS. Effect of data normalization on fuzzy clustering of DNA microarray data. BMC Bioinformatics. 2006;7:134. [Crossref] [PubMed] [PMC]
16. Duyu M, Durmaz B, Gunduz C, Vergin C, Yilmaz Karapinar D, Aksoylar S, et al. Prospective evaluation of whole genome microRNA expression profiling in childhood acute lymphoblastic leukemia. Biomed Res Int. 2014;2014:967585. [Crossref] [PubMed] [PMC]
17. [Link] (Erişim tarihi 14.04.2021)
18. [Link] (Erişim tarihi 14.04.2021)
19. KEGG Pathway Database [Internet]. Wiring diagrams of molecular interactions, reactions and relations. Available from: [Link] (Erişim Tarihi: 14.04.2021)
20. [Link] (Erişim tarihi 14.04.2021)
21. TargetScanHuman [Internet]. Search for predicted microRNA targets in mammals. © 2006-2018 Whitehead Institute for Biomedical Research. Available from: (erişim tarihi: 14.04.2021) [Link]
22. Del Vescovo V, Denti MA. microRNA and Lung Cancer. Adv Exp Med Biol. 2015;889:153-77. [Crossref] [PubMed]
23. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005;15;65(16):7065-70. [Crossref] [PubMed]
24. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A. 2005;8;102(10):3627-32. [Crossref] [PubMed] [PMC]
25. Gironella M, Seux M, Xie MJ, Cano C, Tomasini R, Gommeaux J, et al. Tumor protein 53-induced nuclear protein 1 expression is repressed by miR-155, and its restoration inhibits pancreatic tumor development. PNAS. 2007;104(41):16170-5. https://www.pnas.org/content/104/41/16170 [Crossref] [PubMed] [PMC]

26. Li QJ, Chau J, Ebert PJ, Sylvester G, Min H, Liu G, et al. miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell. 2007;6;129(1):147-61. [PubMed]
27. Ma Z, Qiu X, Wang D, Li Y, Zhang B, Yuan T, et al. MiR-181a-5p inhibits cell proliferation and migration by targeting Kras in non-small cell lung cancer A549 cells. Acta Biochim Biophys Sin (Shanghai). 2015;47(8):630-8. [Crossref] [PubMed]
28. Li Y, Kuscu C, Banach A, Zhang Q, Pulkoski-Gross A, Kim D, et al. miR-181a-5p inhibits cancer cell migration and angiogenesis via downregulation of matrix metalloproteinase-14. Cancer Res. 2015;1;75(13):2674-85. [Crossref] [PubMed] [PMC]
29. He S, Zeng S, Zhou ZW, He ZX, Zhou SF. Hsa-microRNA-181a is a regulator of a number of cancer genes and a biomarker for endometrial carcinoma in patients: a bioinformatic and clinical study and the therapeutic implication. Drug Des Devel Ther. 2015;18;9:1103-75. [Crossref] [PubMed] [PMC]
30. Korhan P, Erdal E, Atabey N. MiR-181a-5p is downregulated in hepatocellular carcinoma and suppresses motility, invasion and branching-morphogenesis by directly targeting c-Met. Biochem Biophys Res Commun. 2014;8;450(4):1304-12. [Crossref] [PubMed]
31. Palamarchuk A, Efanov A, Nazaryan N, Santanam U, Alder H, Rassenti L, et al. 13q14 deletions in CLL involve cooperating tumor suppressors. Blood. 2010;13;115(19):3916-22. [Crossref] [PubMed] [PMC]
32. Pekarsky Y, Croce CM. Role of miR-15/16 in CLL. Cell Death Differ. 2015;22(1):6-11. [Crossref] [PubMed] [PMC]
33. Zanesi N, Balatti V, Bottoni A, Croce CM, Pekarsky Y. Novel insights in molecular mechanisms of CLL. Curr Pharm Des. 2012;18(23):3363-72. [Crossref] [PubMed]
34. Wu J, Ji A, Wang X, Zhu Y, Yu Y, Lin Y, et al. MicroRNA-195-5p, a new regulator of Fra-1, suppresses the migration and invasion of prostate cancer cells. J Transl Med. 2015;13;289:2-15. [Link]
35. Xu H, Hu YW, Zhao JY, Hu XM, Li SF, Wang YC, et al. MicroRNA-195-5p acts as an anti-oncogene by targeting PHF19 in hepatocellular carcinoma. Oncol Rep. 2015;34(1):175-82. [Crossref] [PubMed]
36. Luo Q, Wei C, Li X, Li J, Chen L, Huang Y, et al. MicroRNA-195-5p is a potential diagnostic and therapeutic target for breast cancer. Oncol Rep. 2014;31(3):1096-102. [Crossref] [PubMed] [PMC]
37. Luo Q, Zhang Z, Dai Z, Basnet S, Li S, Xu B, et al. Tumor-suppressive microRNA-195-5p regulates cell growth and inhibits cell cycle by targeting cyclin dependent kinase 8 in colon cancer. Am J Transl Res. 2016;15;8(5):2088-96. [PubMed] [PMC]
38. Zanette DL, Rivadavia F, Molfetta GA, Barbuzano FG, Proto-Siqueira R, Silva WA Jr, et al. miRNA expression profiles in chronic lymphocytic and acute lymphocytic leukemia. Braz J Med Biol Res. 2007;40(11):1435-40. [Crossref] [PubMed]
39. Zhi F, Wang Q, Deng D, Shao N, Wang R, Xue L, et al. MiR-181b-5p downregulates NOVA1 to suppress proliferation, migration and invasion and promote apoptosis in astrocytoma. PLoS One. 2014;9;9(10):e109124. [PubMed] [PMC]
40. Retraction Statement: "Overexpression of miR-708 and its targets in the childhood common precursor B-cell ALL" by Xue Li, MMed, Dong Li, PhD, Yong Zhuang, MMed, Qing Shi, BSc, Wei Wei, MMed, and Xiuli Ju, MD, PhD. Pediatr Blood Cancer. 2017;64(5). [Crossref] [PubMed]
41. Uzhga-Rebrov O, Kuleshova G. Problems of fuzzy clustering of microarray data. IT and Management Science. 2010;44(1):51-4. [Crossref]