

RESEARCH

Open Access



# Discovery of inflammatory bowel disease-associated miRNAs using a novel bipartite clustering approach

Md. Altaf-UI-Amin<sup>1\*</sup>, Mohammad Bozlul Karim<sup>1</sup>, Pingzhao Hu<sup>2</sup>, Naoaki ONO<sup>1</sup> and Shigehiko Kanaya<sup>1</sup>

From Joint 30th International Conference on Genome Informatics (GIW) & Australian Bioinformatics and Computational Biology Society (ABACBS) Annual Conference  
Sydney, Australia. 9–11 December 2019

## Abstract

**Background:** Multidimensional data mining from an integrated environment of different data sources is frequently performed in computational system biology. The molecular mechanism from the analysis of a complex network of gene-miRNA can aid to diagnosis and treatment of associated diseases.

**Methods:** In this work, we mainly focus on finding inflammatory bowel disease (IBD) associated microRNAs (miRNAs) by biclustering the miRNA-target interactions aided by known IBD risk genes and their associated miRNAs collected from several sources. We rank different miRNAs by attributing to the dataset size and connectivity of IBD associated genes in the miRNA regulatory modules from biclusters. We search the association of some top-ranking miRNAs to IBD related diseases. We also search the network of discovered miRNAs to different diseases and evaluate the similarity of those diseases to IBD.

**Results:** According to different literature, our results show the significance of top-ranking miRNA to IBD or related diseases. The ratio analysis supports our ranking method where the top 20 miRNA has approximately tenfold attachment to IBD genes. From disease-associated miRNA network analysis we found that 71% of different diseases attached to those miRNAs show more than 0.75 similarity scores to IBD.

**Conclusion:** We successfully identify some miRNAs related to IBD where the scoring formula and disease-associated network analysis show the significance of our method. This method can be a promising approach for isolating miRNAs for similar types of diseases.

**Keywords:** IBD, BiClusO, MRM (miRNA regulatory module), MTI (miRNA target interaction)

## Background

Inflammatory bowel disease (IBD) results in disorders in different digestive organ parts with prolonged pain and disruption. The specific causes of IBD, including ulcerative colitis (UC) and Crohn's disease (CD), remain unknown. Epidemiology of IBD shows the increasing rate by every year where prevention or cure of this disease is still intractable [1]. Even different risk factors,

such as ethnicity, smoking, age, family history and gender, are attributed to the IBD, scientists are trying to find other evidence by analyzing the IBD related genomic data. Recently it has been discovered that non-coding RNAs (ncRNAs) are able to control gene expression in a sequence specific manner. Among various types of ncRNAs, MicroRNAs (miRNAs) appear as important cytoplasmic regulators of gene expression. miRNAs are non-coding RNAs of the approximated length of 22 nucleotides, playing important roles in gene splicing and post-transcriptional regulation of gene. Recent studies revealed that there is a strong connection between the

\*Correspondence: [amin-m@is.naist.jp](mailto:amin-m@is.naist.jp)

<sup>1</sup>Nara Institute of Science and Technology, Ikoma 630-0192, Japan  
Full list of author information is available at the end of the article



regulatory mechanism of miRNA and disease etiology [2–4]. As an example, Overexpression of miR-21 was found in mice to contract pre-B malignant lymphoid-like phenotype tumors. Complete tumor regression is achieved by inactivating this miRNA [5]. Therapeutic inhibition of miRNAs using antisense oligomers (called anti-miRs) has also been shown to reduce tumor growth [6]. Scientists have developed miRNA - target interaction (MTIs) databases based on different proven scientific methods which can be used to drill down the functional modules of specific miRNA sets and their target interactions. Previously we developed methods to identify the IBD associated genes from the integrated analysis of transcriptome data and protein-protein interactions (from HIPPIE database). We also compared our results with three different databases namely HuGENet, DisGeNet, CTD and another genome wide association study (GWAS) with respective IBD genes of 849, 866, 129 and 335. Finally we identified a group of IBD related genes with different confidence scores [7]. A miRNA-regulatory module (MRM) is a subset of MTIs where groups of miRNAs participate cooperatively by regulating a bunch of genes to control different biological processes [8]. The MTIs can be represented as a bipartite graph. A bipartite graph is a network of two disjoint sets of nodes where each edge connects a node from one set to a node from the other set. No edge is allowed within any single set. A bicluster is a high density (in terms of connected edges) subgraph of a bipartite graph. There are various applications of biclustering in different fields of study. In biology, gene expression under certain conditions forms a bipartite network which helps to identify the cellular response, disease diagnosis and pathway analysis. Biological network analysis of the pairwise combinations of protein, miRNA, metabolite, conserved functional subsequences, and factor binding sites can predict or understand different cellular mechanisms. Graph convolutional and deep learning methods are also popular technique on prioritizing or predicting the outcome of a gene or disease from such network [9–11]. In the current work, we mainly focused on MRM detection from MTIs by a new biclustering approach we recently developed [12, 13]. We then searched the IBD related genes in MRMs detected in MTI networks. We evaluated the relevance of the miRNAs with IBD by counting their occurrences in different MRMs and their interactions with known IBD genes. Finally, we normalized the score of each miRNA for different MTIs database and evaluated the importance of different miRNA.

## Methods

### IBD gene set

We previously proposed a method for predicting IBD risk genes based on currently known IBD risk genes collected from DisGeNet database and differentially expressed

genes determined using gene expression data [7]. In that work we created a disease relevant Protein-Protein Interaction (PPI) network by selecting data from Human Integrated Protein-Protein Interaction reference (HIPPIE) database and then determined high density clusters in the PPI network utilizing DPCLUSO algorithm [14–16]. Finally, from the statistically significant clusters, we determined 909 genes as potential IBD genes as our novel predictions. We also downloaded IBD related genes from other online sources such as HuGENet [17], Comparative Toxicogenomics Database (CTD) [18], DisGeNet [19] and literatures related to genome wide association study (GWAS) [20–23]. By combining all data, we created a set of IBD related genes comprising 2245 genes.

### miRNA-mRNA/Gene Interaction dataset

miRNAs act as post-transcriptional regulators of the target messenger RNAs (mRNAs) via degradation and/or translational repression. Each miRNA can be linked to a gene. There are databases that have accumulated information of interactions between miRNAs and their target mRNAs/genes. We have collected such information from four different online databases as follows: mirWalk (<http://mirwalk.umm.uni-heidelberg.de/>) [24], DIANA (<http://diana.imis.athena-innovation.gr/>) [25], miRecords (<http://c1.accurascience.com>) [26], miRTarbase (<http://mirtarbase.mbc.nctu.edu.tw>) [27]. Sometimes in a database there are multiple entries of the same miRNA-mRNA interacting pairs which are different in terms of other attributes such as tissue sample, binding probability, binding site position, cell line, tissue, disease category etc. Hence, we used below criteria to select the interactions from these databases. For mirWalk, each mRNA-miRNA interaction has at least 2 evidences (duplicate in the database) and for DIANA, it has at least 4 evidences since more evidences identified for a given mRNA-miRNA show the interaction has higher quality. For miRTarbase, we selected the interaction with the term 'Functional MTI' since they have higher quality than 'Non-functional MTI' tagged interactions. Table 1 shows the number of interactions, and associated number of miRNAs and genes we selected for the current study from the four different databases.

### MRM extraction

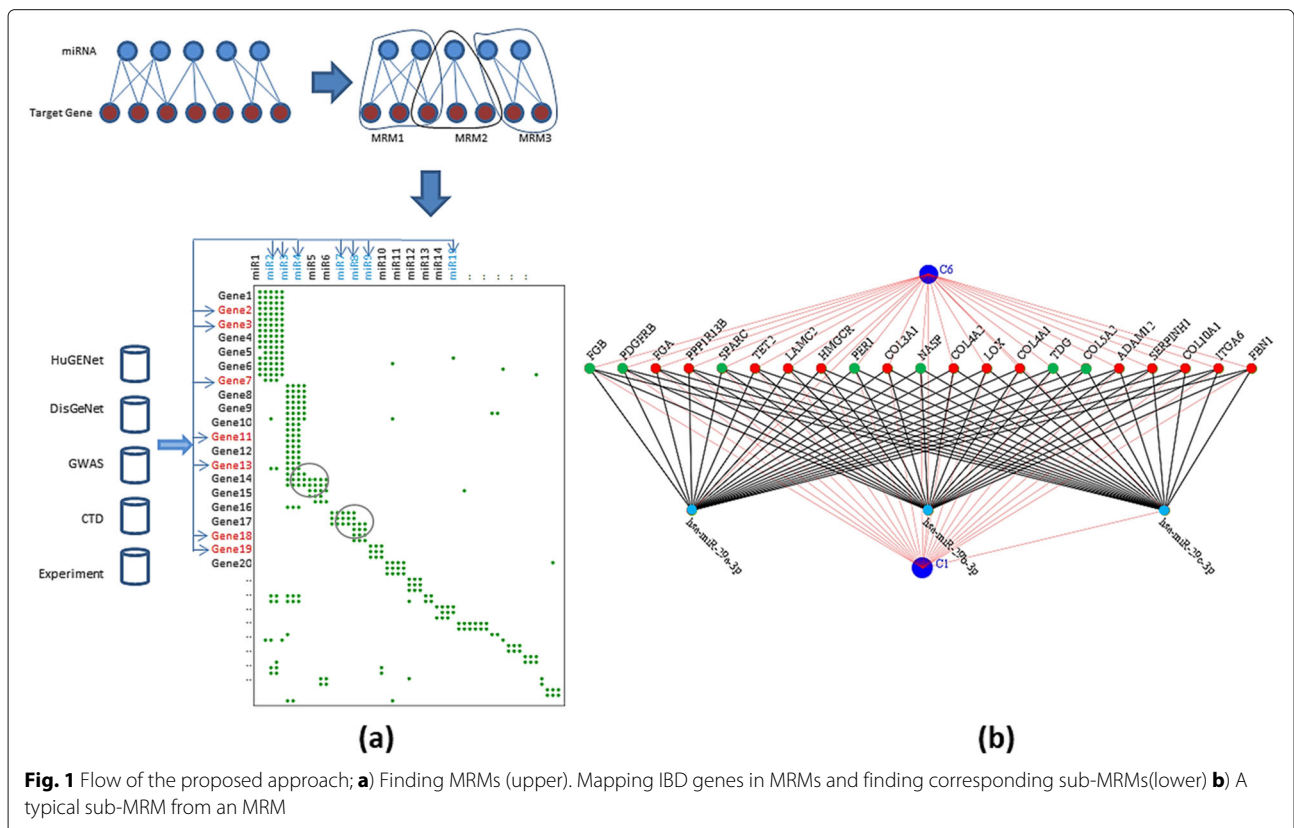
The interactions between miRNAs and mRNAs can be represented as a bipartite graph which is called miRNA - target interaction (MTI) network. A bipartite graph is a network of two disjoint sets of nodes where each edge connects a node from one set to a node from the other set. No edge is allowed within any one set. A bicluster is a high density (in terms of connected edges) subgraph of a bipartite graph. In an MTI bipartite network, the miRNAs are a set of nodes and mRNAs are the other set of nodes.

The biclusters in an MTI are called miRNA-regulatory modules (MRM). We recently developed a biclustering algorithm called BiClusO [12, 13]. This algorithm was mainly developed for identifying biclusters from a bipartite graph as the miRNA-mRNA network we used in this study. Since a given miRNA can bind to different sets of mRNAs, which implies a given miRNA can be found in different MRMs or biclusters. Based on our algorithm, the bicluster set from a bipartite graph can be overlapped to a certain degree i.e any node may belong to more than one cluster. This is an inherent property of the bipartite graph. The basic theory of BiClusO is to convert a two dimensional problem to one dimensional one by data folding, solve it by one-dimensional algorithm and unfold it again. Thus, the BiClusO algorithm first converts the bipartite graph to a simple graph by taking any node set and measuring the association between those node pairs using relation number and Tanimoto coefficient, then performs simple graph clustering using the polynomial-time heuristic algorithm DPCLUSO we developed before [14]. Finally, the attachment of the nodes from the second set creates each bicluster. Figure 1a shows the flow of extracting MRMs from a MTI network by BiClusO. Two examples of overlapping between biclusters are indicated by circles in the lower part of Fig. 1a. BiClusO algorithm generates a reasonable number of overlapping biclusters under the optimized parameter settings [7, 12]. In the current

work for BiClusO we utilized the following parameter setting: cluster density=0.5, cluster property=0.5, relation number=3, Tanimoto coefficient =0.33 and attachment probability =0.5. Each bicluster is called an MRM. A typical MRM is constructed by a set of miRNAs that are strongly connected with a set of genes. An MRM contains system level information on relations between miRNAs and genes. From the MRMs we created IBD related sub-MRMs by identifying the presence of IBD genes. As mentioned above we selected 2245 IBD genes from different databases and studies. For each bicluster, these genes were matched and corresponding miRNAs were separated. Thus IBD related sub-MRMs were generated. Figure 1b shows a typical sub-MRM from an MRM. The green-colored nodes in the gene side are IBD genes. Red-colored nodes indicate the non IBD genes and aqua colored nodes are the attached miRNA in this MRM. The blue nodes attached by thin red edges are overlapping MRMs to this MRM. Usually, the total number of sub-MRMs is less than the number of MRMs and the size of each sub-MRM is less than the size of the corresponding MRM.

**Relevance Score Calculation**

We generated IBD related sub-MRMs from 4 different MTIs (as mentioned in Table 1) separately. Within each set of sub-MRMs, we calculated the relevance score of



**Fig. 1** Flow of the proposed approach; **a**) Finding MRMs (upper). Mapping IBD genes in MRMs and finding corresponding sub-MRMs(lower) **b**) A typical sub-MRM from an MRM

**Table 1** Number of interactions, miRNA and mRNA on different datasets

Dataset	Interactions	miRNA	Gene
mirWalk	17,290	51	4621
mirTarbase	8157	735	2756
miRecords	1710	249	1097
DIANA	17,535	521	5491

individual miRNA as a measure of its relation with IBD by using the following formula

$$RS_{miRNA(i)} = NoofIBD_{miRNA(i)} * Noofcluster_{miRNA(i)} \tag{1}$$

Here

$RS_{miRNA(i)}$  = Relevance score of  $i^{th}$  miRNA

$NoofIBD_{miRNA(i)}$  = number of IBD genes attached to  $i^{th}$  miRNA in the IBD MRM set

$Noofcluster_{miRNA(i)}$  = number of IBD MRMs attached to  $i^{th}$  miRNA

The relevance score signifies the attachment of each miRNA to different sub-MRMs and IBD genes. The total number of interactions, miRNAs, and genes are different among the datasets used in our experiment. We downloaded the latest updated versions of the datasets and observed that the collecting method and importance of each interaction might vary in terms of attributes and parameters in different data sources. Interactions included in more than one dataset might be more accurate and important. Therefore, for a combined ranking of the miRNAs in terms of their relevance to IBD genes is needed. After finding the miRNA sets and their corresponding relevance score from the four different datasets, all sets were merged. We normalized the score of individual miRNA in each dataset and proposed an overall score for each miRNA as follows

$$TRSmRNA_i = \sum_{n=1}^4 \frac{RSn_i}{C_n} \sum_{n=1}^4 En_i \tag{2}$$

Here

$TRSmRNA_i$  is the total relevance score of  $i^{th}$  miRNA based on all dataset

$RSn_i$  is the relevance score of  $i^{th}$  miRNA in  $n^{th}$  dataset

$C_n$  the number of cluster in  $n^{th}$  dataset

$En_i$  is the Boolean value measuring whether  $i^{th}$  miRNA is in the  $n^{th}$  dataset

As an example, hsa-let-7b-5p was found in three datasets: DIANA, mirTarbase, and mirWalk. The total number of biclusters generated from DIANA, mirTarbase,

and mirWalk datasets are 650, 64 and 1579. In DIANA hsa-let-7b-5p was attached to 44 biclusters with 48 IBD genes. In mirTarbase it was attached with 1 bicluster with 1 IBD gene and in mirWalk, it was attached with 209 biclusters with 69 IBD genes. So the relevance score for this miRNA in three datasets are 2112, 1, and 14421. The total score is  $(2112/650+1/64+14421/1579)(1+1+1) = 37.21$

### Results and discussion

miRNA isolating is frequently used in the diagnosis and monitoring of different diseases. Numerous studies have identified miRNAs as a potential biomarker for different diseases. Different databases of miRNA and mRNA interaction are created by compiling experimental results of different studies. Most of the databases have varied attributes with different quality along with miRNA and mRNA. Researchers are updating their databases by collecting the biological and medicinal experimental results. One of the biggest challenges in this work was to select the valid interactions which have strong evidence on the basis of other attribute values. Important attributes and frequency of reported interactions were considered in the selection process and duplicate or triplicate miRNA-mRNA pairs were removed from the final dataset utilized in this study.

### Ranking of the miRNAs

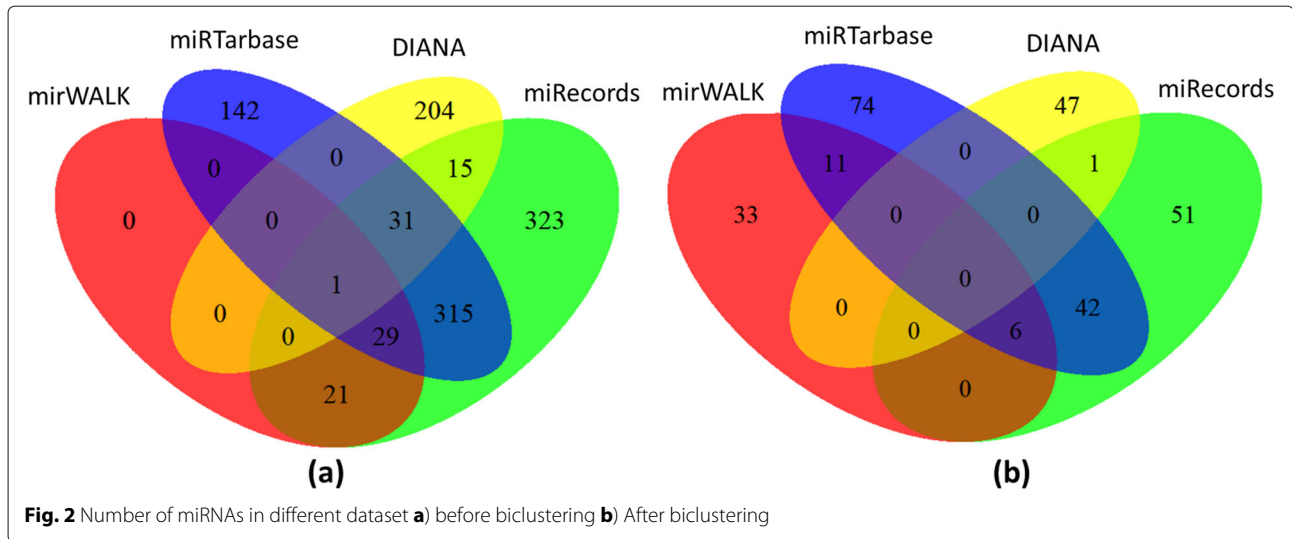
Based on miRWalk dataset, we generated 1579 biclusters from which we found 1011 sub-MRMs encompassing 50 miRNAs and 333 genes. Top 10 miRNAs according to relevance score are hsa-let-7d-5p, hsa-let-7a-5p, hsa-let-7e-5p, hsa-let-7c-5p, hsa-let-7b-5p, hsa-miR-106a-5p, hsa-miR-106b-5p, hsa-let-7f-5p, hsa-let-7i-5p.

64 biclusters were generated from mirTarbase dataset out of which we found 41 IBD related sub-MRMs encompassing 100 miRNAs and 128 genes. Top 10 miRNAs are hsa-miR-221-3p, hsa-miR-29b-3p, hsa-miR-222-3p, hsa-miR-34c-5p, hsa-miR-200c-3p, hsa-miR-29c-3p, hsa-miR-200b-3p, hsa-miR-29a-3p, hsa-miR-34b-3p, hsa-miR-24-3p.

23 biclusters were generated from small dataset miRecords where 20 sub-MRMs with 48 miRNAs and 54 genes were found. Out of them, top 10 miRNAs are hsa-miR-16, hsa-miR-15a, hsa-miR-17, hsa-miR-29a, hsa-miR-181a, hsa-miR-29b, hsa-miR-1, hsa-miR-221, hsa-miR-20a, hsa-miR-34b.

DIANA dataset produced 650 biclusters with 423 sub-MRMs where 133 miRNAs and 340 genes were found. Top 10 miRNAs in this dataset are hsa-miR-1-3p, hsa-miR-16-5p, hsa-miR-15a-5p, hsa-miR-15b-5p, hsa-miR-124-3p, hsa-miR-103a-3p, hsa-miR-27a-3p, hsa-miR-107, hsa-miR-20a-5p, hsa-let-7b-5p. Venn diagram from Fig. 2 shows the number of miRNAs in different datasets before





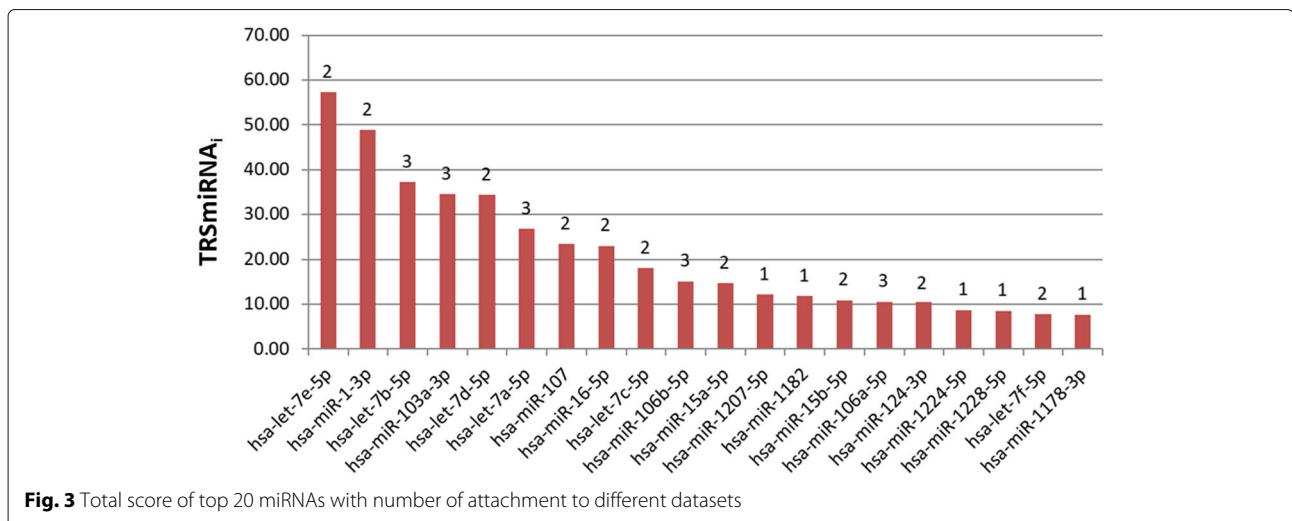
biclustering (a) and after detection of sub-MRMs (b). From the Venn diagram of Fig. 2a, it is evident that most of the datasets have an almost distinct set of miRNAs. The 265 miRNAs included in Fig. 2b are IBD related miRNAs detected by our approach. Out of them 6 are common in three different datasets and 53 are common miRNAs on two different datasets. Figure 3 shows the top 20 miRNA according to the total relevance score where 15 of them were found in at least two datasets.

**Relevance of the top miRNAs to IBD**

We also evaluated the ratio of IBD related genes associated with the top 20 selected miRNAs to those of all of the 265 selected miRNAs and it appears that IBD related genes are enriched in the top 20 miRNAs. The total number of IBD

genes attached to the top 20 miRNAs is 493 whereas total number of IBD genes attached to all 265 miRNAs is 664. Thus an approximate ratio of 10:1 is achieved in terms of attachment to the IBD genes for the top 20 miRNAs.

Out of the top 20 selected miRNAs, 6 miRNAs are related to the miR-7 family. miR-7 family has 9 members which are let-7a, let-7b, let-7c, let-7d, let-7e, let-7f, let-7g, let-7i, and miR-98. miRNAs related to this family regulate various biological functions such as cell proliferation, cell cycle, stem cell biology, metabolism, and migration, progression, and chemoresistance. miR-7 is downregulated on different types of cancer such as colon cancer [28], gastric tumors [29] etc. Patients with inflammatory bowel disease (IBD) are at significantly increased risk of colorectal cancer (CRC) [30–32], principally resulting



from the pro-neoplastic effects of chronic intestinal inflammation [33].

Using imperfect base pairing to the 3'-UTR the mature let-7 negatively regulates the expression of target mRNAs at a posttranslational level [34]. The expression levels of let-7 microRNAs in stem and progenitor cells are maintained low during the normal development process. Expression levels increase when the progenitor cells differentiate [35]. The downregulation of let-7 promotes migration and invasion of normal intestinal epithelial cells and CRC cells [36]. The downregulation of let-7 or upregulation of either LIN28A or LIN28B has been reported to be related to the prognosis in CRC patients in critical stage. The expression level of LIN28B was inversely correlated to that of mature let-7a in human CRC [36]. From an experiment, 38% out of 600 CRC patients were found to be highly expressed of LIN28A or LIN28B [37]. Let-7 microRNAs are also downregulated in different types of cancers such as hepatocellular carcinoma (HCC), gastric adenocarcinoma, pancreatic cancer, ovarian cancer, prostate cancer, Burkitt lymphoma, renal cell carcinoma, breast cancer, and melanoma [38].

Expression of hsa-let-7e-5p is markedly upregulated in HHM RC. Subsequent assessment of the expression of hsa-let-7e-5p target genes implicated that it may be a prognostic biomarker for RC with HHM [39]. Both inflamed and non-inflamed terminal ileal mucosa in adult patients with active CD have their distinct miRNA expression patterns compared with healthy controls for hsa-let-7b-5p [40]. let-7d has a significant impact on epithelial-to-mesenchymal transition (EMT) and formation of cancer initiating cells which are resistant to irradiation and chemical exposure and responsible for cancer metastasis [41]. In patients with stage II CRC hsa-miR-103a-3p is reported as a promising predictive biomarkers for tumor recurrence [42]. Expression of miR-16 is elevated in CD and UC peripheral blood [43]. Overexpression of miR-106b-5p suppress the CRC cell migration and inhibits the invasion and metastasis of colorectal cancer by targeting CTSA [44].

Upregulation of miR-15a-5p in IBD patient is reported in [45]. By sponging miR-1207-5p a long noncoding RNA BC032469 upregulates hTERT expression which promotes proliferation in gastric cancer [46]. There was a significant negative correlation between miR-1182 and hTERT which attenuates gastric cancer. miR-15b-5p is down-regulated in CRC cells and tissues. The inhibitory effects of miR-15b-5p on cell apoptosis and enhancement of drug sensitivity are mediated by the down-regulation of its NF- $\kappa$ B1 and IKK- $\alpha$  targets [47]. Long non-coding RNA FER1L4 exerts tumor suppressive effects on colon cancer by mediating miR-106a-5p repression [48]. The expression level of miR-106a is elevated in Intestinal biopsy, peripheral blood/serum cell of UC and CD patient [49].

miR-124 is downregulated by regulating STAT3 expression in colon tissues of pediatric patients with UC [50]. The expression level of miR-124-3p is increased in the advanced stage of CRC patients. miR-124-3p works as a tumor suppressor gene in astrocytomas by targeting the repression of protein PIM1 [51]. miR-1224-5p has the colitogenic ability in the gut epithelium and is directly associated with IBD disease [52]. miR-1228 is downregulated in gastric cancer tissues also overexpression of miR-1228 significantly inhibited the proliferation and colony formation of gastric cancer cells [53].

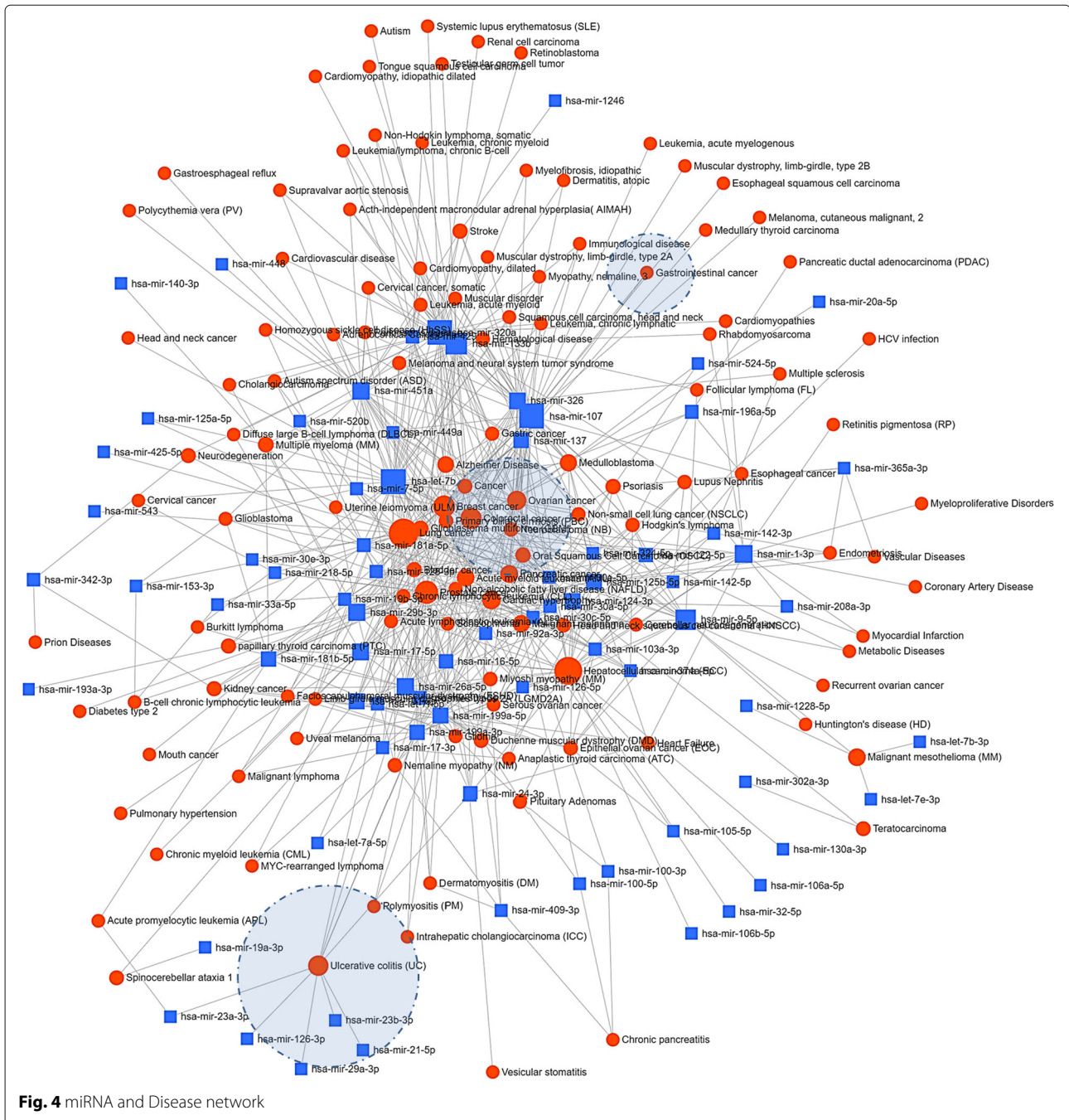
#### miRNA disease network

We used 265 our identified miRNAs to miRNet (<https://www.mirnet.ca>) and got the associated miRNA-disease network [54]. 74 out of the 265 miRNAs are included in this unweighted network with a minimum degree cutoff = 1. The network is shown in Fig. 4 where the circular nodes correspond to diseases and the rectangular nodes are miRNAs. The sizes of the nodes are proportional to their respective degrees. The layout of Fig. 4 expresses the centralization of the nodes with higher degree hence the association of a disease with a significant number of miRNAs are plotted in the center.

The dotted circles are the highlighted region of three IBD associated diseases i.e. Ulcerative colitis, CRC and Gastrointestinal Cancer. From the network, most of the associations are with different type of cancer e.g Lung cancer, hepatocellular carcinoma (HCC), prostate cancer, breast cancer, colorectal cancer, ovarian cancer, and pancreatic cancer. We also identified ulcerative colitis with 10 different miRNAs and gastrointestinal cancer with two different miRNAs. We found 126 different types of diseases on this network where 20% of them are associated with inflammation. Some of these diseases are Acute lymphoblastic leukemia (ALL), Chronic pancreatitis, Dermatitis, atopic, Dermatomyositis (DM), Multiple sclerosis, Polymyositis (PM), Psoriasis, etc. This also implies that inflammatory pathway analysis related to those diseases can be applied for IBD where inflammation is a common symptom.

#### Disease Similarity

We also evaluated the similarity of identified disease set from the miRNA-disease network to Inflammatory bowel disease. We used DisGeNET to find those disease and their corresponding gene set. The database contains gene-disease associations from UNIPROT, CGI, ClinGen, Genomics England, CTD (human subset), PsyGeNET, and Orphanet. We used 'DOSE' package of R and evaluated the disease similarity by using cluster-sim function. Figure 5 shows the similarity of identified disease set with the Inflammatory bowel disease. Matching score 1 indicates maximum similarity. The results



show that 71% diseases are similar to IBD with a score more than 0.75. Therefore, it can be concluded that our approach is a promising method for prioritizing IBD related miRNAs and this method can be applied to other diseases.

**Conclusion**

Dysregulation of single or multiple miRNAs can affect normal cellular function i.e proliferation, metabolism,

apoptosis, cell cycle, stem cell division, neuronal gene expression which are the major cause of different diseases in human. In recent years scientists have proved the regulation of miRNAs to turn the cancer cell as malignant. Rapid discovery and progress of different clinical experiment accumulate the MRM data that can aid to discover the molecular mechanism of disease development. Our present goal was to narrow down the large domain of the multidimensional database and discover effective

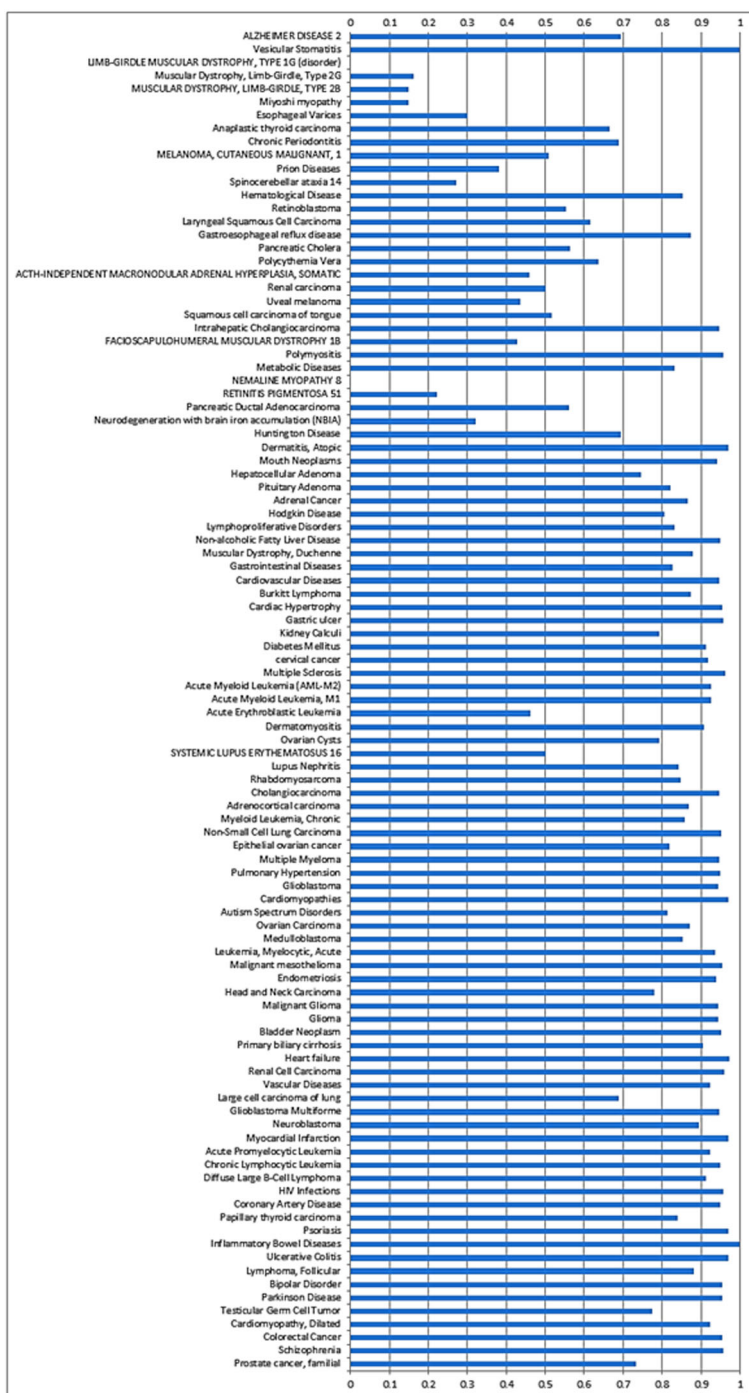


Fig. 5 Disease similarity between IBD and different diseases

information. In this work, we successfully identified some important IBD related miRNAs. We also searched the literature for the association of these miRNAs to UC, CD and similar types of disease like CRC and Gastrointestinal cancer and provided evidences in support of our predictions.

**Abbreviations**

ALL: Acute lymphoblastic leukemia; BiClusO: Biclustering algorithm with overlapping; CD: Crohn's disease CD; CRC: Colorectal cancer; CTD: Comparative toxicogenomics database; DM: Dermatomyositis; DIANA: A database of miRNA:gene interactions; DisGeNet: Disease gene network; DPCLusO: Simple cluster algorithm with overlapping; GWAS: genome wide association study; HCC: hepatocellular carcinoma; HIPPIE: Human integrated



protein-protein interaction reference; HuGENet: The human genome epidemiology network; IBD: Inflammatory bowel disease; MRM: miRNA regulatory module; mRNA: Messenger RNA; miRNA: micro RNA; miRecords: A resource for animal miRNA-target interactions; miRTarBase: A curated database of microRNA-target interactions; mirWalk: A comprehensive atlas of predicted and validated miRNA-target interactions; MTI: miRNA target interaction; ncRNA: Non-coding RNA; PM: Polymyositis; PPI: Protein-protein interaction; RS: Relevance score; TRS: Total relevance score; UC: Ulcerative colitis

#### Acknowledgements

We thank Md Zobaer Hasan, Ph.D of Rohto pharmaceuticals for his critical insight into biological phenomena.

#### About this supplement

This article has been published as part of BMC Medical Genomics Volume 13 Supplement 3, 2020: Proceedings of the Joint International GIW & ABACBS-2019 Conference: medical genomics (part 2). The full contents of the supplement are available online at <https://bmcmcdgenomics.biomedcentral.com/articles/supplements/volume-13-supplement-3>.

#### Authors' contributions

Md. A-U-A and MBK designed the research and conducted the experiments. SK,PH, NO guided the research with valuable comments. All authors have read and approved the final manuscript.

#### Funding

This work was supported by NAIST Global Collaborative Program 2019 and partially supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (16K07223 and 17K00406), NAIST Big Data Project and by Research Manitoba, Health Sciences Centre Foundation and Mitacs of Canada. Publication costs have been funded by NAIST Global Collaborative Program 2019. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

Biclustering tool BiClusO is available in the link "<http://www.knapsackfamily.com/BiClusO/>". Data is available on request from the corresponding author.

#### Ethics approval and consent to participate

Not Applicable

#### Consent for publication

Not Applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Nara Institute of Science and Technology, Ikoma 630-0192, Japan. <sup>2</sup>University of Manitoba, Winnipeg, Canada.

Received: 22 November 2019 Accepted: 7 January 2020

Published: 24 February 2020

#### References

- Gasparetto M, Guariso G. Highlights in IBD epidemiology and its natural history in the paediatric age. *Gastroenterol Res Pract*. 2013;2013.
- Vishnoi A, Rani S. MiRNA biogenesis and regulation of diseases: an overview. In: *MicroRNA Profiling*. New York: Humana Press; 2017. p. 1–10.
- Gommans WM, Berezikov E. Controlling miRNA regulation in disease. In: *Next-Generation MicroRNA Expression Profiling Technology*. Humana Press; 2012. p. 1–18.
- Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. *Development*. 2005;132(21):4653–62.
- Medina PP, Nolde M, Slack FJ. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature*. 2010;467(7311):86–90.
- Cheng CJ, Bahal R, Babar IA, et al. MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. *Nature*. 2015;518(7537):107–10.
- Eguchi R, et al. An integrative network-based approach to identify novel disease genes and pathways: a case study in the context of inflammatory bowel disease. *BMC Bioinformatics*. 2018;19:264.
- Walsh CJ, Hu P, Batt J, dos Santos CC. Discovering MicroRNA-Regulatory Modules in Multi-Dimensional Cancer Genomic Data: A Survey of Computational Methods. *Cancer Informat*. 2016:68–73.
- Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods*. 2019;166:4–11.
- Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457–66.
- Li Y, Kuwahara H, Yang P, Song L, Gao X. PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks. *bioRxiv*. 2019:532226.
- Karim MB, Huang M, Naoaki ONO, Kanaya S, Amin MAU. BiClusO: A novel biclustering approach and its application to species-VOC relational data. *IEEE/ACM transactions on computational biology and bioinformatics*. 2019. <https://doi.org/10.1109/TCBB.2019.2914901>.
- Karim MB, Kanaya S, Amin MAU. Comparison of BiClusO with Five Different Biclustering Algorithms Using Biological and Synthetic Data. *International Conference on Complex Networks and their Applications*. Cham: Springer; 2018.
- Amin MAU, Wada M, Kanaya S. Partitioning a PPI network into overlapping modules constrained by high-density and periphery tracking. *ISRN Biomathematics*. 2012;2012:11.
- Amin MAU, Shinbo Y, Mihara K, Kurokawa K, Kanaya S. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics*. 2006;7(1):207.
- Amin MAU, Tsuji H, Kurokawa K, Asahi H, Shinbo Y, Kanaya S. DPCLus: a density-periphery based graph clustering software mainly focused on detection of protein complexes in interaction networks. *J Comput Aided Chem*. 2006;7:150–6.
- Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet*. 2008;40(2):124–5. <https://doi.org/10.1038/ng0208--124>.
- Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, Mcmorran R, Wieggers J, Wieggers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res*. 2017;45. <https://doi.org/10.1093/nar/gkw838>.
- Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database: J Biol Databases Curation*. 2015;2015:028. <https://doi.org/10.1093/database/bav028>.
- Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979–86. <https://doi.org/10.1038/ng.3359>.
- Anderson CA, Boucher G, Lees CW, Franke A, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43(3):246–52. <https://doi.org/10.1038/ng.764>.
- Franke A, McGovern DPB, Barrett JC, Wang K, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42(12):1118–25. <https://doi.org/10.1038/ng.717>.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008;40(8):955–62. <https://doi.org/10.1038/ng.175>.
- Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS ONE*. 2018;13(10).
- Vlachos IS, et al. DIANA miRPath v. 2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res*. 2012;40(W1):W498–504.
- Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res*. 2009;37:D105–10.
- Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*. 2017;46(D1):D296–302.

28. Akao Y, Nakagawa Y, Naoe T. let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol Pharm Bull*. 2006;29(5):903–6.
29. Zhang HH, Wang XJ, Li GX, Yang E, Yang NM. Detection of let-7a microRNA by real-time PCR in gastric carcinoma. *World J Gastroenterol*. 2007;13(20):2883.
30. Rutter MD, Saunders BP, Wilkinson KH, Rumbles S, Schofield G, Kamm MA, Williams CB, Price AB, Talbot IC, Forbes A. Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis. *Gastroenterology*. 2006;130(4):1030–8.
31. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*. 2001;48(4):526–35.
32. Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. *Gastroenterology*. 2011;140(6):1807–16.
33. Stidham RW, Higgins PDR. Colorectal Cancer in Inflammatory Bowel Disease. *Clin Colon Rectal Surg*. 2018;31(03):168–178. <https://doi.org/10.1055/s-0037-1602237>.
34. Berindan-Neagoe I, Calin GA. Molecular pathways: microRNAs, cancer cells, and microenvironment. *Clin Cancer Res*. 2014;20(24):6247–53.
35. Viswanathan SR, Daley GQ. Lin28: A microRNA regulator with a macro role. *Cell*. 2010;140(4):445–9.
36. King CE, Wang L, Winograd R, Madison BB, Mongroo PS, Johnstone CN, Rustgi AK. LIN28B fosters colon cancer migration, invasion and transformation through let-7-dependent and-independent mechanisms. *Oncogene*. 2011;30(40):4185.
37. Tu H-C, Schwitalla S, Qian Z, LaPier GS, Yermalovich A, Yuan-Chieh K, Chen S-C, et al. LIN28 cooperates with WNT signaling to drive invasive intestinal and colorectal adenocarcinoma in mice and humans. *Genes Dev*. 2015;29(10):1074–86.
38. Wang T, Wang G, Hao D, Xi L, Wang D, Ning N, Li X. Aberrant regulation of the LIN28A/LIN28B and let-7 loop in human malignant tumors and its effects on the hallmarks of cancer. *Mol Cancer*. 2015;14(1):125.
39. Chen W, Lin G, Yao Y, Chen J, Shui H, Yang Q, Wang X, et al. MicroRNA hsa-let-7e-5p as a potential prognosis marker for rectal carcinoma with liver metastases. *Oncol Lett*. 2018;15(5):6913–24.
40. Guo Z, Rong W, Gong J, Zhu W, Yi L, Wang Z, Li N, Li J. Altered micro RNA expression in inflamed and non-inflamed terminal ileal mucosa of adult patients with active Crohn's disease. *J Gastroenterol Hepatol*. 2015;30(1):109–16.
41. Kolenda T, Przybyla W, Teresiak A, Mackiewicz A, Lamperska KM. The mystery of let-7d—a small RNA with great power. *Contemp Oncol*. 2014;18(5):293.
42. Zhang J-X, Song W, Chen Z-H, Wei J-H, Liao Y-J, Lei J, Ming H, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*. 2013;14(13):1295–306.
43. Wu F, Zhang S, Dassopoulos T, Harris ML, Bayless TM, Meltzer SJ, Brant SR, Kwon JH. Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflamm Bowel Dis*. 2010;16(10):1729–38.
44. Ni S, Weng W, Midie X, Wang Q, Tan C, Sun H, Wang L, Huang D, Xiang D, Sheng W. miR-106b-5p inhibits the invasion and metastasis of colorectal cancer by targeting CTSA. *OncoTargets Ther*. 2018;11:3835.
45. Honardoost MA, Naghavian R, Ahmadinejad F, Hosseini A, Ghaedi K. Integrative computational mRNA-miRNA interaction analyses of the autoimmune-deregulated miRNAs and well-known Th17 differentiation regulators: An attempt to discover new potential miRNAs involved in Th17 differentiation. *Gene*. 2015;572(2):153–62.
46. Lü MH, Tang B, Zeng S, Hu CJ, Xie R, Wu YY, Wang SM, He FT, Yang SM. Long noncoding RNA BC032469, a novel competing endogenous RNA, upregulates hTERT expression by sponging miR-1207-5p and promotes proliferation in gastric cancer. *Oncogene*. 2016;35(27):3524.
47. Zhao C, Qi Z, Zhang C, Wang G, Yao Y, Huang X, Zhan F, et al. miR-15b-5p resensitizes colon cancer cells to 5-fluorouracil by promoting apoptosis via the NF- $\kappa$ B/XIAP axis. *Sci Rep*. 2017;7(1):4194.
48. Yue B, Bo S, Liu C, Zhao S, Zhang D, Fudong Y, Yan D. Long non-coding RNA Fer-1-like protein 4 suppresses oncogenesis and exhibits prognostic value by associating with miR-106a-5p in colon cancer. *Cancer Sci*. 2015;106(10):1323–32.
49. Omidbakhsh A, et al. Micro-RNAs -106a and -362-3p in Peripheral Blood of Inflammatory Bowel Disease Patients. *Open Biochem J*. 2018;12:78–86. <https://doi.org/10.2174/1874091X01812010078>.
50. Koukos G, Polyarchou C, Kaplan JL, Morley-Fletcher A, Gras-Miralles B, Kokkotou E, Baril-Dore M, Pothoulakis C, Winter HS, Iliopoulos D. MicroRNA-124 regulates STAT3 expression and is down-regulated in colon tissues of pediatric patients with ulcerative colitis. *Gastroenterology*. 2013;145(4):842–52.
51. Deng D, Wang L, Chen Y, Li B, Xue L, Shao N, Wang Q, Xia X, Yang Y, Zhi F. MicroRNA-124-3p regulates cell proliferation, invasion, apoptosis, and bioenergetics by targeting PIM1 in astrocytoma. *Cancer Sci*. 2016;107(7):899–907.
52. Lee J, Park EJ, Kiyono H. MicroRNA-orchestrated pathophysiologic control in gut homeostasis and inflammation. *BMB Rep*. 2016;49(5):263.
53. Jia L, Chen J, Xie C, Shao L, Zhipeng X, Zhang L. microRNA-1228\* impairs the pro-angiogenic activity of gastric cancer cells by targeting macrophage migration inhibitory factor. *Life Sci*. 2017;180:9–16.
54. Fan Y, Xia J. miRNet—Functional Analysis and Visual Exploration of miRNA-Target Interactions in a Network Context. In: *Computational Cell Biology*. New York: Humana Press; 2018. p. 215–33.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

