Database, 2021, 1–20
doi:10.1093/database/baab031
Original article

Original article

emiRIT: a text-mining-based resource for microRNA information

Debarati Roychowdhury[1,*], Samir Gupta[1], Xihan Qin[2], Cecilia N. Arighi[2] and K. Vijay-Shanker[1]

[1]Department of Computer and Information Sciences, University of Delaware, 101 Smith Hall, 18 Amstel Ave, Newark, DE 19716, USA and [2]Department of Computer and Information Sciences, Center of Bioinformatics and Computational Biology, University of Delaware, 15 Innovation Way, Room 205, Newark, DE 19711, USA

*Corresponding author: Tel: +1 302-831-2711; Fax: 302-831-8458; Email: droyc@udel.edu

Citation details: Roychowdhury, D., Gupta, S., Qin, X. *et al.* emiRIT: a text-mining-based resource for microRNA information. *Database* (2021) Vol. 2021: article ID baab031; doi:10.1093/database/baab031

Received 23 November 2020; Revised 15 March 2021; Accepted 4 May 2021

Abstract

microRNAs (miRNAs) are essential gene regulators, and their dysregulation often leads to diseases. Easy access to miRNA information is crucial for interpreting generated experimental data, connecting facts across publications and developing new hypotheses built on previous knowledge. Here, we present extracting miRNA Information from Text (emiRIT), a text-miningbased resource, which presents miRNA information mined from the literature through a user-friendly interface. We collected 149 ,233 miRNA –PubMed ID pairs from Medline between January 1997 and May 2020. emiRIT currently contains 'miRNA –gene regulation' (69 ,152 relations), 'miRNA disease (cancer)' (12 ,300 relations), 'miRNA –biological process and pathways' (23, 390 relations) and circulatory 'miRNAs in extracellular locations' (3782 relations). Biological entities and their relation to miRNAs were extracted from Medline abstracts using publicly available and in-house developed text-mining tools, and the entities were normalized to facilitate querying and integration. We built a database and an interface to store and access the integrated data, respectively. We provide an up-to-date and user-friendly resource to facilitate access to comprehensive miRNA information from the literature on a large scale, enabling users to navigate through different roles of miRNA and examine them in a context specific to their information needs. To assess our resource's information coverage, we have conducted two case studies focusing on the target and differential expression information of miRNAs in the context of cancer and a third case study to assess the usage of emiRIT in the curation of miRNA information.

Database URL: https://research.bioinformatics.udel.edu/emirit/

Introduction

microRNAs (miRNAs) are non-coding small RNAs that regulate gene expression at the post-transcriptional level. The majority of protein-coding genes are controlled by miRNAs, suggesting that most biological processes are subjected to miRNA-dependent regulation (1). Several studies have also shown miRNA implications in cancer and neurodegenerative diseases (2–8). Experimental findings regarding miRNAs, covering various contents such as target information, differential expression of miRNAs and their role in diseases, are scattered across multiple publications and databases (9, 10). For example, consider a biomedical researcher interested in knowing the miRNAs that are differentially expressed in the context of triple-negative breast cancer. The researcher may be interested in the biological processes impacted by such miRNAs or their target genes, specifically if they are mentioned in this disease context. Such information may be found in multiple publications and different databases. However, conducting a literature survey to extract all these related information is time-consuming and a laborious process and requires significant switching between different resources. Another significant issue is that miRNA-based publications have been growing exponentially (Figure 1), making it difficult for existing miRNA resources to be up to date.

As such, there is a critical need for resources that can significantly reduce the cumbersome information retrieval process and quickly obtain relevant and integrated information from miRNA-related studies. For this reason, we designed extracting miRNA Information from Text (emiRIT), which mines different miRNA information from PubMed abstracts published between January 1997 and May 2020 (11). Our resource combines the mined results of several text-mining tools on a large scale in one place with a unified output format. Thus, our resource offers a variety of miRNA information in one location, making the navigation between different information easier. To include the most widely studied miRNA aspects, we focus on recognizing biological entities (bioentities) of type (i) 'miRNA', (ii) 'gene', (iii) 'disease (currently only cancer)', (iv) 'biological processes and pathways' and (v) 'extracellular locations' (transporters and biofluids). Bioentities are linked to publicly available standard ontologies/databases to ensure smooth querying, sorting and filtering capabilities in our interface and expand querying abilities to integrate with external resources. The

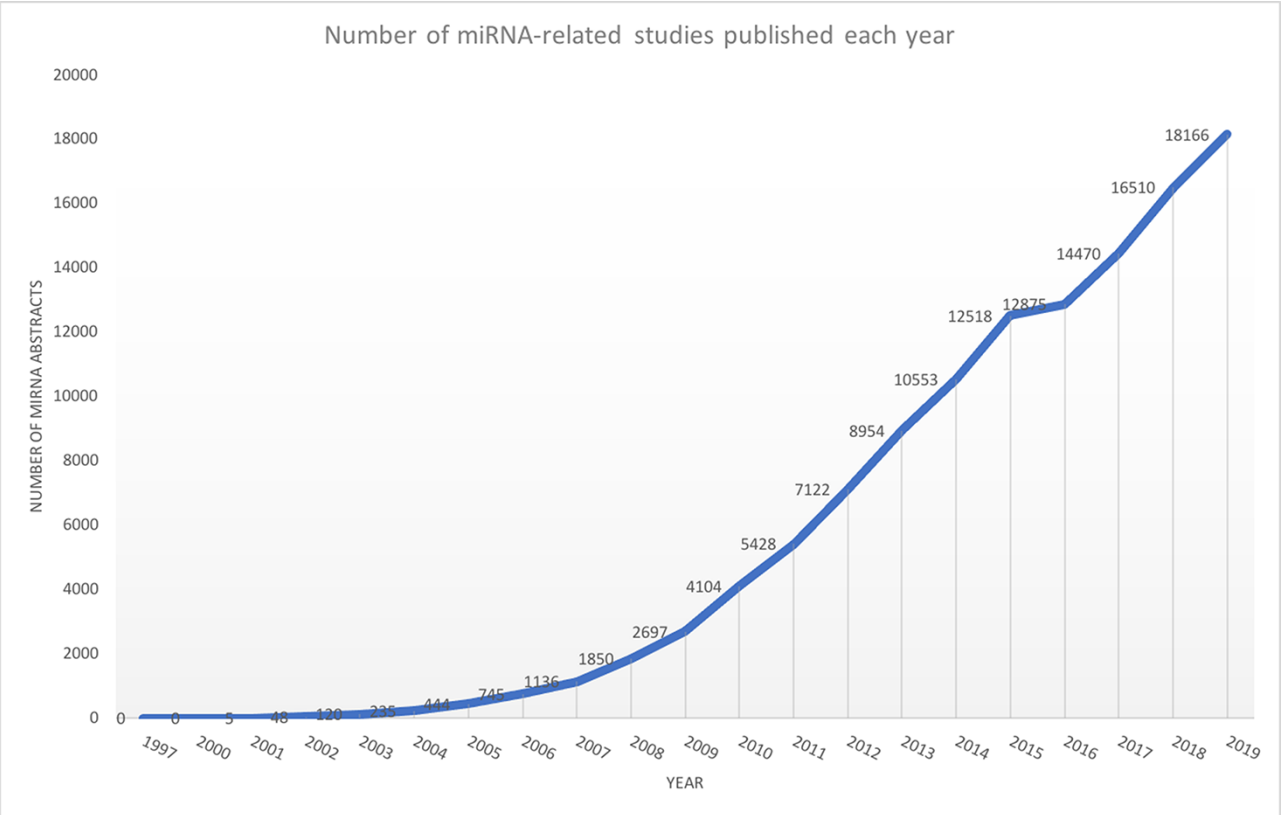


Figure 1. Exponential growth of miRNA publications obtained from Medline using keyword 'miRNA' OR 'microRNA'.

text-mining tools are applied simultaneously on the miRNA literature to provide consistent and regular updates from new papers. Since all the information are mined from abstracts, we are able to link all our extracted results to the literature. Currently, many of the existing resources do not have direct link to the literature evidence, hindering a full interpretation of the miRNA information due to lack of context.

A unique aspect of emiRIT is that it provides a more detailed picture of an miRNA's role in the context of a disease. The different roles of miRNAs we detect in diseases are illustrated by the following examples:

(i) Role in disease process and outcome: 'High miR-21 expression is associated with poor survival and poor therapeutic outcome.' [PMID: 18230780 (12)]
(ii) Role in disease treatment: 'The role of miR-181a in conferring cellular resistance to radiation treatment was validated both in cell culture models and in mouse tumor xenograft models.' [PMID: 22847611 (13)]
(iii) Role as biomarker: 'Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma' [PMID: 19179615 (14)]
(iv) Role as therapeutic target: 'These findings suggest that miR-24 could be an effective drug target for treatment of hormone-insensitive prostate cancer or other types of cancers.' [PMID: 20195546 (15)]
(v) Unspecified role in disease: 'Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer.' [PMID: 18196926 (16)]
(vi) Differential expression in disease: 'Extrapolation of this study to human primary HCCs revealed that miR-122 expression was significantly ($P = 0.013$) reduced in 10 out of 20 tumors compared to the pair-matched control tissues.' [PMID: 16924677 (17)]

The mined information is stored in a database and presented to users through an interface. On emiRIT's interface, users can examine different miRNA aspects in one place, smoothly navigate between various aspects for a broader understanding of miRNAs' role and also narrow down the information to a specific biological context. Our main goal is to provide an up to date and user-friendly resource to facilitate access to relevant miRNA information from the literature. In the remainder of this paper, we discuss related work, followed by the description of the pipeline to extract, store and present miRNA information comprehensively mined from Medline abstracts.

Related work

As discussed in the previous section, miRNA data are scattered across multiple publications and databases. Several of the databases shown in Table 1 are literature based and curated by experts, which makes these resources high quality but hard to maintain and keep up with the most recent results. The last update of several of these databases in the table dates back to more than 5 years ago. Recent efforts on miRNA annotations come from Intact (18) and the GO consortium (19) and have focused on specific topics, such as rare diseases and cardiovascular and neurodegenerative diseases, respectively.

Due to the sheer amount of miRNA publications, as indicated in Figure 1, text-mining methods have been increasingly adopted for automatic extraction of relations between an miRNA and a target/process/disease to assist in tasks such as database curation and knowledge discovery [miRNEST (34), miRSel (35), miRTex (36), miRCancer (37), miRiaD (38), Murray *et al.*, 2010 (39), DES-ncRNA (40)]. Except for a few, most of the approaches are limited by their narrow scope. DES-ncRNA uses only simple co-occurrences within sentences to find miRNA connections but lack the robustness to capture connections beyond co-occurrence. Work done by Murray *et al.*, 2010 is network oriented. They do not directly find an miRNA's connection to a disease or process. Instead, they combine an miRNA's connection to genes and then use curated databases to get a gene's connection to diseases and processes to finally generate a Cytoscape network.

emiRIT, on the other hand, focuses on extracting connections between an miRNA and a gene/disease/process/extracellular location by utilizing text-mining tools that capture patterns from the syntactic structure of sentences. The following section describes the pipeline for development of emiRIT.

System design

This section describes the design and structure of the two major components: a database to store relevant miRNA information and an interface to interact with the stored data.

emiRIT database

Database content

To meet the desired needs of our resource, the database stores the following:

(i) Various relations involving miRNA that are extracted from text in publications.
(ii) The biological context, such as a disease context or a process context, in which the above-extracted relations were mentioned to provide more perspective to specific user needs.
(iii) The literature evidence for these miRNA relations so that users can read the sentence or the whole abstract to interpret the information and get a

Table 1. A sample of existing literature-based miRNA databases

Category	Resource name	Short description	Year of last update
miRNA target	Intact (18)		2021
	GO (19)		2021
	DIANA-TarBase (20)	Validated miRNA–target interactions	2018
	miRWalk (21)		2020
	miRTarBase (22)		2018
	miRecords (23)		2013
miRNA-transcription factor	TransMir (24)	Validated transcription factor miRNA regulations	2018
miRNA disease	miR2Disease (25)	Validated dysregulated miRNAs in human disease	2009
	OncomirDB (26)	Validated or potentially pathogenic roles of dysregulated miRNAs in cancer	2014
	HMDD (27)	Experimentally supported human miRNA–disease associations	2019
miRNA pathway/process	miRwayDB (28)	Validated miRNA–pathway associations	2020
	GO Biological Process (29)		2021
miRNA location	GO cellular component (29)	Validated miRNA-subcellular location annotations	2021
miRNA-extracellular locations	miRandola (30)	Validated extracellular circulating non-coding RNAs	2017
miRNA expression	dbDEMC (31)	Differentially expressed miRNAs in human cancers	2017
miRNA expression	PhenoMir (32)	Manually curated database collecting differentially regulated miRNA expression in diseases and biological processes	2011
miRNA sequence, miRNA expression	miRBase (33)	Collects and constructs information about published miRNA sequences and expression profiles	2018
	miRNEST (34)	Combines sequence, expression from external database with predicted targets, mirtrons and miRNA gene structure	2016

complete understanding of the relations in a fuller context.

(iv) Standardized/normalized text mentions of all entities allowing for extended querying capabilities and smoother integration with external resources and ontologies leading to a more connected resource. Standardizing and normalizing entities using pertinent ontologies will also provide access to additional descriptions for these entities.

Creation of the database

In this section, we will describe the processing of text from scientific publications as well as discuss the recognition and normalization of entities and the extraction of the relations between them. Figure 2 shows the workflow of how the database is created using abstracts and viewed in the interface.

Text preprocessing To support the extraction of the relations and provide literature evidence, we store miRNA-related abstracts from PubMed split into sentences in the database. We download abstracts mentioning miRNAs from Medline using the query 'miR OR miRNA OR microRNA' on PubMed. We split each abstract into sentences using the Stanford CoreNLP sentence splitter (41).

Entity recognition and normalization As discussed earlier, we will focus on the following types of bioentities in our miRNA resource: miRNAs, genes, diseases, biological processes and pathways, and extracellular locations (transporters and biofluids). For miRNA detection, we use a regular-expression-based in-house tool (see Supplementary Table S1 for the regular expressions). The detected mentions are normalized to the corresponding family ID in miRBase (33). For genes and disease mentions, we use PubTator

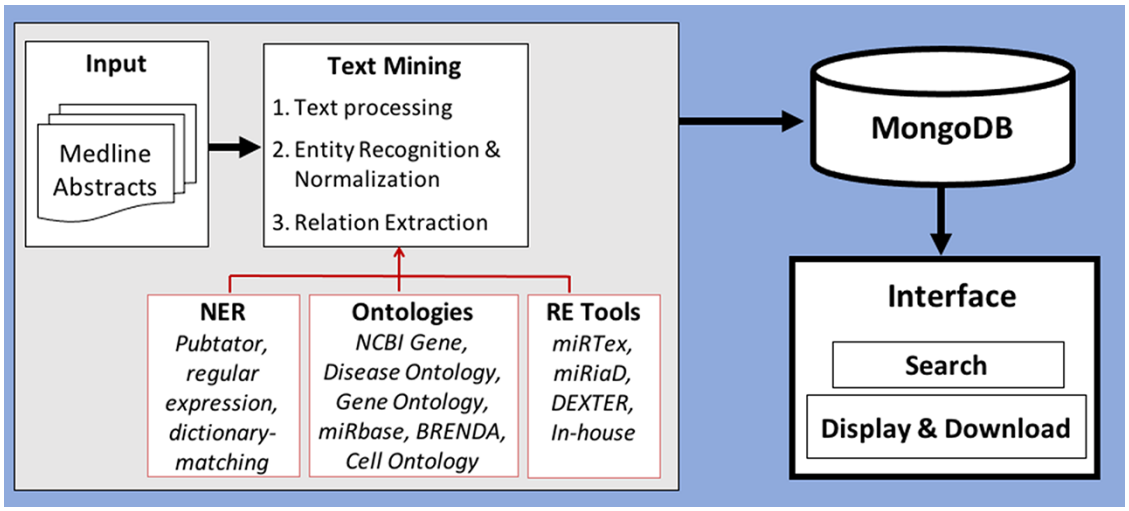


Figure 2. Workflow of creation of database by processing and storing miRNA-relevant information and viewing through an interface.

(42) to detect the mentions. PubTator normalizes genes to NCBI Gene IDs and diseases to MESH IDs. We keep the same normalization for genes, but we map these MESH IDs to Disease Ontology IDs (DOID) (43) using the publicly available mapping table from http://purl.obolibrary.org/obo/doid.obo.

For all other entities, we develop our own dictionary-based matching technique to obtain the closest match, as follows. For biological processes/pathways, the dictionary is created using a combination of terms and their synonyms from the cellular processes of the biological process branch of gene ontology (GO) (29), the pathway ontology (PW) (44) and terms mined from text using patterns. For example, on encountering text such as 'cellular processes such as migration, invasion and cell death', we extract the three listed process terms, commonly appearing terms mined from large amounts of Medline abstracts were thus collected. The process terms detected are normalized to GO and PW based on exact string matching. For extracellular locations, we use exRNA forms and fluid samples from miRandola (30) to build our dictionary and normalize the fluid mentions from the text using Brenda tissue ontology (45).

Relation extraction This subsection discusses the extraction of various relations stored in the database:

miRNA-gene

One of the main relations we capture is between an miRNA and a gene since miRNAs are important gene regulators. To capture such a relation, we use a text-mining tool called miRTex (36), which detects the miRNA and its target gene (a gene regulation of an miRNA). miRTex also detects where an miRNA regulates gene expression—either indirectly or when it is not clear if the regulation is a direct result of targeting. In other words, miRTex detects three types of relations—(i) miRNA and gene (when a direct relation cannot be inferred in the sentence, (ii) miRNA and target (when a direct relation is detected and (iii) gene and miRNA (for relations about regulators of miRNA expression.

Example 1: 'Mechanistic studies disclosed that, miR-340 over-expression suppressed several oncogenes including p-AKT, EZH2, EGFR, BMI1 and XIAP' [PMID: 25831237 (46)]

Example 2: 'Furthermore, ROCK1 was validated as a direct functional target miR-340 and silencing of ROCK1 phenocopied the anti-tumor effect of mR-340' [PMID: 25831237 (46)]

Example 3: 'TGF-β1 increased miR-34a expression in cardiac fibroblasts' [PMID: 25322725 (47)]

miRTex has been evaluated and found to be a robust extraction system of the three different types of miRNA–gene relations from abstracts and full-length articles with high precision, recall and F-scores close to 0.9.

miRNA-process

The next relation we focus on is between an miRNA and a process since most biological processes are subjected to miRNA-dependent regulation. For this purpose, we have extended another text-mining tool, miRiaD (38). Central to miRiaD is the detection of the 'involvement', 'regulation' and 'association' between an miRNA and a disease aspect. For our resource, we extend the connections with association, involvement and regulation (CAIR) framework of miRiaD to make the connections between miRNAs and processes and pathways, irrespective of the presence of disease terms in the abstract. In the example below, the CAIR

framework will detect that miR-29b positively regulates the apoptotic process.

Example 4: microRNA-29b promotes high-fat diet-stimulated endothelial permeability and apoptosis in apoE knock-out mice by down-regulating MT1 expression. [PMID: 25131924 (48)]

miRNA-disease

As mentioned before, the varied roles of miRNA in diseases have been widely researched, including their role as potential biomarkers and therapeutic targets, their impact on the treatment of diseases and disease outcomes. Instead of just stating that there is an association between an miRNA and a disease, we seek to present a more detailed picture of the role of an miRNA in context of disease by distinguishing between these different roles. Specifically, these roles are (i) impact of miRNA on disease process and outcome, (ii) influence on disease treatment, (iii) diagnostic role as biomarkers, (iv) role as therapeutic targets in diseases and (v) others, where the particular role is not clear, but the miRNA is associated with a disease or regulates a disease.

Since miRiaD was developed to capture the different ways an miRNA is linked to a disease or a disease aspect, we extend miRiaD significantly for our purpose. For the first two relation types, we use the CAIR framework from miRiaD, whereas for the next two relations, we use a group of rules clubbed together to form the 'is_a' framework that can capture relations such as 'X is a Y', or 'X acts as Y' or 'X serves as Y'. While miRiaD clubbed all the five different roles together and called them 'disease aspects', we have enhanced miRiaD's ability to take the arguments of the relations and we separate the disease aspects based on the type of the arguments. As an illustration of argument-based separation, examples 5 and 6 below show how a sentence depicting an miRNA's role as biomarker and therapeutic target is structured.

Example 5: From a clinical point of view, our study emphasizes miR-122 as a diagnostic and prognostic marker for HCC progression.' [PMID: 19617899 (49)]

Example 6: 'Our data suggest that miR-429 may serve as a potential anticancer target for the treatment of HCC'. [PMID: 2844423 (50)]

In example 5, the 'is_a' framework captures 'miR-122' is_a 'diagnostic and prognostic marker', while the same rule captures 'miR-429' is_a 'anticancer target for the treatment' in example 6. In both cases, we look at the type of the arguments of the relation and separate them into 'miRNA is a biomarker' and 'miRNA is a therapeutic target'. miRiaD had reported a high recall and precision with an F-score close to 0.90 when evaluated on a curation task as well as for general extraction of miRNA to disease associations.

Finally, knowing which miRNAs are differentially expressed in disease is important, especially for understanding or generating hypotheses about the underlying causes. To capture the up- or down-regulation expression of miRNAs in disease vs non-disease states from the research literature, we use a tool called DEXTER (51). DEXTER was designed after an extensive study of textual mentions of comparisons (52). It detects the differential expression levels as well as the location of the expression levels such as in cell lines or tissue samples, patient groups, control and others. DEXTER was evaluated and precision greater than 0.90 with an F-score close to 0.80 was reported for general extraction of differentially expressed genes and miRNAs in the context of diseases. In the section discussing view of miRNA-aspects in the interface, Figure 8 shows how the interface will display the different miRNA roles extracted using miRiaD and DEXTER to cater to the information needs of our resource. Since both miRiaD and DEXTER were developed specifically for cancer, we have currently restricted emiRIT to only cancer with plans to extend to other diseases in future.

miRNA-extracellular locations

miRNAs are increasingly being studied as potential biomarkers of diseases because of their abundance and stability in extracellular fluids, transported via membrane-bound vesicles such as exosomes or complexed with high-density lipoprotein (53, 54). Hence, we focus on extracting information about miRNAs in extracellular locations. We use the extended dependency graph (EDG) (55) framework to capture the syntactic structure of sentences and extract direct relations between an miRNA and biofluids, such as tear, serum, plasma and others or extracellular transporter forms, such as vesicles, exosomes, protein complexes and others. As a first step, we use the EDG framework to capture simple patterns that are focused on high precision, for cases where an miRNA and an extracellular location appear in close textual proximity. For the second step, we focus on capturing cases where the miRNA and the extracellular location are not in close textual proximity, but the extracellular location is explicitly mentioned in the experimental context of the paper. Examples 7 and 8 show the two different types of cases we focus on.

Example 7: 'After qRT-PCR validation, only one seminal plasma miRNA, let-7b-5p, was found significantly decreased in severe asthenozoospermia cases compared with healthy controls.' [PMID: 29653228 (56)]

Example 8: [PMID: 32373058 (57)]

(i) We analyzed the expression of three microRNAs in serum of 18 patients (DMD 13, BMD 5) and 13 controls using droplet digital PCR. [Sentence 2]

(ii) We found that levels of miR-30 c and miR-206 remained significantly elevated in DMD patients relative to controls over the entire study length. [Sentence 7]

For the second step, we implement the patient context (PC) sentence detection from eGARD (58), which provide information about the patients involved in the study. Our assumption, following the study of at least 20 abstracts, is based on the fact that any extracellular fluid samples from these patients are highly likely to have a connection to the miRNAs being explored in the same paper. Using this new relation extraction tool, we were able to extract 3782 miRNA–extracellular location pairs using the EDG framework and 2173 pairs using the PC sentences. We sampled about 100 abstracts from our database that contained mentions of miRNA and extracellular locations. From these 100 abstracts, we were able to find 136 miRNA-extracellular location pairs using both the EDG framework and PC sentences. We manually checked each of these 136 pairs and found 133 of them were indeed correctly paired, whereas three of them were paired incorrectly. We plan to improve our new miRNA–extracellular location relation extraction tool in future and conduct further evaluation by comparing the results from the tool to a manually annotated dataset.

Database structure

Based on our previous experience of providing access to information stored in a database to users via an interface in iTextMine (59), which is an integrative text-mining system for knowledge extraction developed in our lab, we choose to store our data using a standardized JSON format (60). This format is a lightweight data-interchange text-based format commonly used for transmitting data in web applications. Our data are centered around miRNA-relevant abstracts that undergo various text processings to retrieve the entities and their relations within each abstract. These document-centric data are then stored in a non-relational database. We use MongoDB (61) since it can accommodate diverse types of data, including documents, and the abstracts can be easily represented using a JSON format and then directly inserted into the database as a document collection. Figure 3 shows an example of how an abstract is represented in the database. The 1 and m in the figure indicate a 1 to many relations. In other words, each document has many sentences, many entities and many relations but has only one title. An example of the JSON format of a document is provided in Supplementary Figure S2. The pipeline for creating the database also includes an 'Update' step, which will ensure that the database is brought up to date every few months and the most current miRNA information is captured in our resource.

emiRIT interface

We created the interface https://research.bioinformatics.udel.edu/emirit/using Flask 1.0.3 and Bootstrap 4.0. The various functionalities of the interface are described in the following subsections.

Search mode

There are two main modes of querying the database through this interface. The first type is the miRNA-centric search, where users can observe the different connections between a specific miRNA stored in the database and other entities. The second type is a context-centric search, where users can observe the different miRNA connections in a

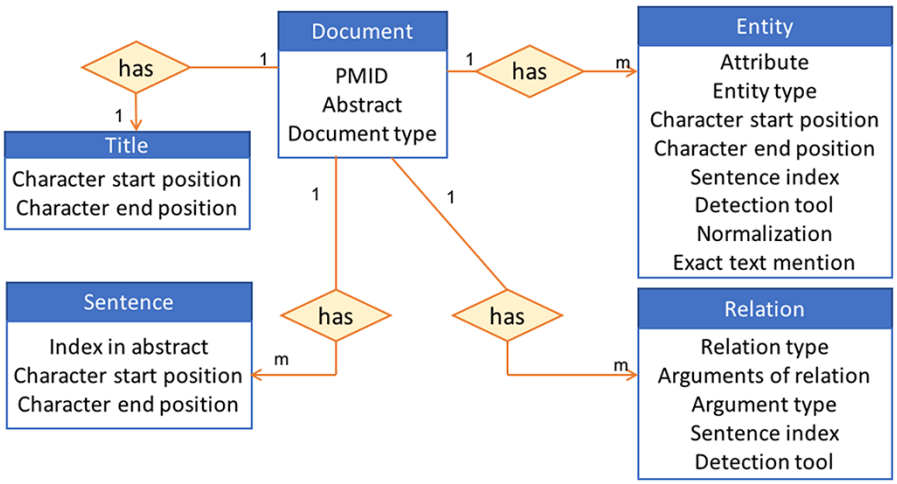


Figure 3. High-level view of the information stored in the database for an abstract.

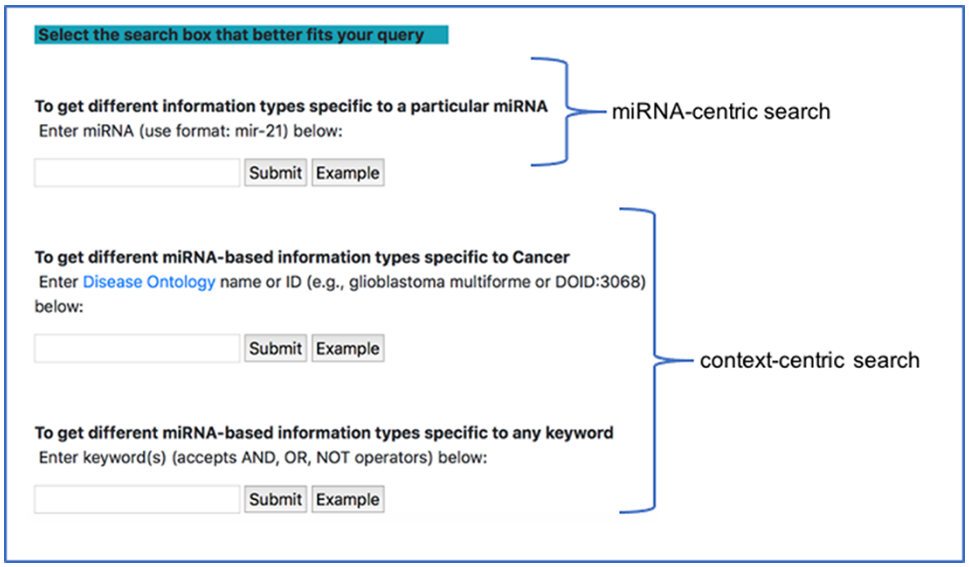


Figure 4. miRNA-centric search and context-centric search mode in the interface.

certain context using a general keyword query. Unlike, the miRNA-centric search, the context-centric search is similar to a PubMed search. Here, we can include separators like AND/OR in our query. To every query, we explicitly add 'miR' or 'miRNA' or 'microRNA' and use the resultant query to search in the NCBI PubMed database and retrieve a list of PMIDs. We then search our database using this list of PMIDs and extract the various information for the common PMIDs from our database.

Since most of the miRNA research is conducted in the context of a disease, we provide a specialized context-centric search, which limits the context to a specific disease. Users can search the database using the DOID or official DOID name and observe the different connections between miRNA and other entities in the context of the corresponding disease. The DOID is used to retrieve the disease name and its synonyms and using a combination of these disease terms and miRNA terms, a query, similar to the general context-centric search, is constructed to search in the NCBI PubMed database. Currently, we narrow our query to only cancer-specific diseases. Additionally, if a user starts typing a disease name in the search box, a drop down of disease names from disease ontology, generated using the NCBO BioPortal widget, is also provided. Figure 4 shows a screenshot of the different search modes in our interface.

View of miRNA aspects

High-level view The resulting page for any of the search queries shows how many documents were retrieved and how many entities were found. A table summarizing the different entities in each document is displayed at the bottom of the page (as shown in Figure 5). This table shows a high-level view of genes, diseases and processes involved in a relation with miRNAs for each document referred by the corresponding PMID. These entities were found to be in a relation with miRNAs by the relation extraction tools we have discussed before. From the search result page shown in Figure 5, users can either navigate to specific aspects of miRNAs using the tabs 'Gene Regulation', 'Biological Process and Pathway', 'Disease' and 'Extracellular location of circulatory miRNA', or they can navigate to a specific document, as described in the following subsections.

Aspect-specific view The aspect-oriented information can be viewed by exploring the tabs at the top of the search result page (refer to Figure 6). For instance, if the user wants to know what different genes are targeted by the different miRNAs in the context of GBM, they can choose the specific 'Gene Regulation' tab (Figure 6). The resulting page (refer to Figure 7) will show the miRNA to target, miRNA to gene and gene to miRNA relations, in the context of GBM, along with the PMID of the abstract as literature evidence. The information displayed from clicking the gene regulation tab does not change the list of genes presented in the previous page. Instead, the new page provides the additional distinction between the three types of miRNA–Gene relations as well as the normalization Ids of miRNAs and genes to expand the knowledge to descriptions of each entity.

The 'Disease' tab (Figure 6) will take us to another page that separates the different disease-oriented information, as shown in Figure 8. As discussed previously, an important component of our work is the ability to distinguish between the different roles or aspects of miRNAs in the context of a disease. Users can look at different aspects, such

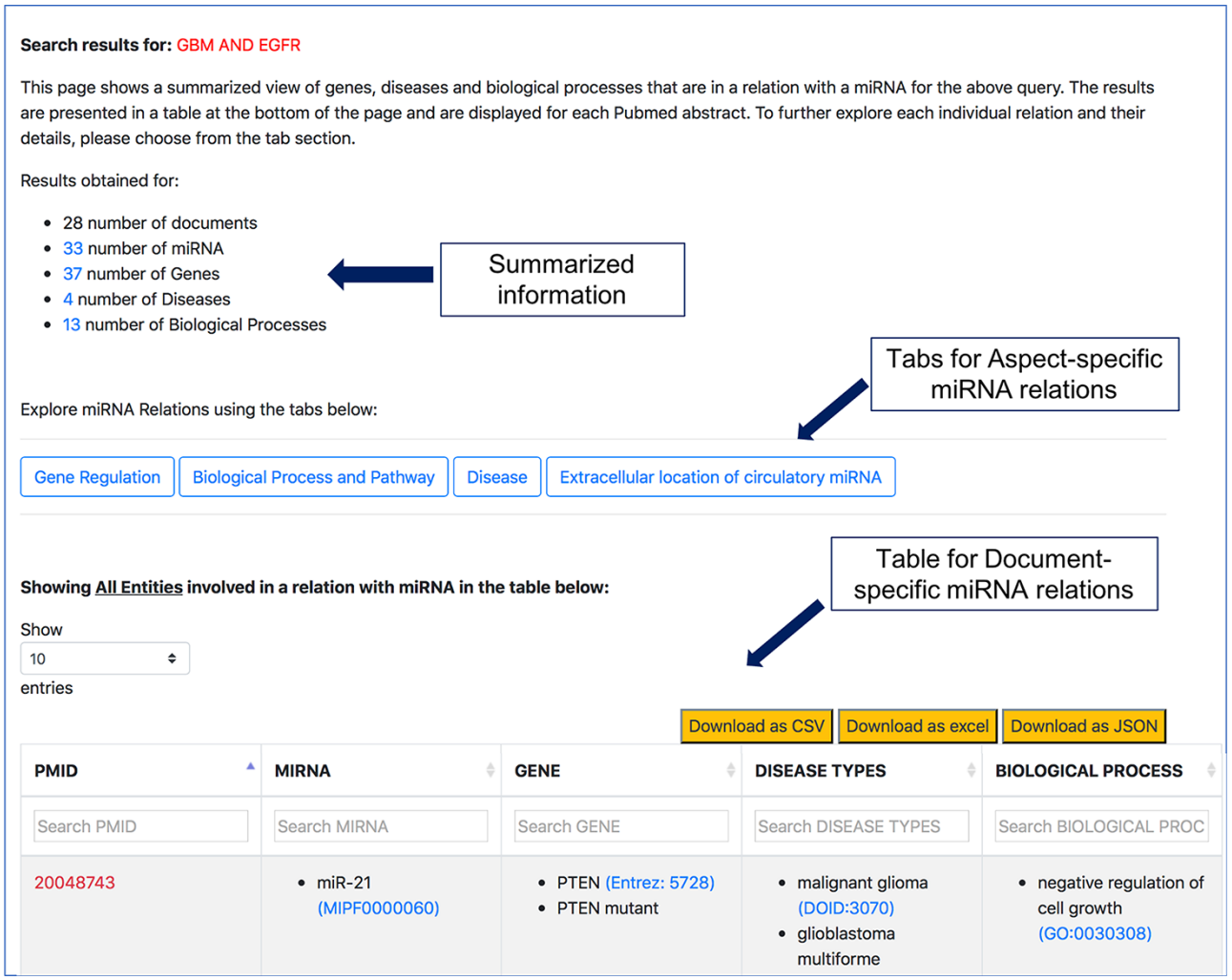


Figure 5. Screenshot of response page for context-centric query 'GBM AND EGFR'.

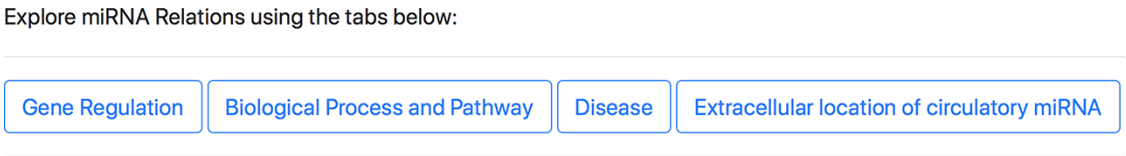


Figure 6. Specific miRNA relation tabs at the top of the response page for a query.

as which miRNAs are up-regulated or down-regulated in disease, what is an miRNA's impact on the outcome of a disease or a disease process, which miRNAs were found to be potential biomarkers and unspecified role in a disease indicated as 'others'. The disease type information in the table specifies the name of the disease associated with the miRNA in a particular abstract.

Document-specific view On clicking the PMID from any of the above pages, users will be taken to a page showing all relations about a single document. Currently, our resource only looks at abstracts since the tools we use are limited to abstracts, but we plan to extend the resource to PubMed Central (PMC) open access papers in future. As shown in Figure 9, this page shows the abstract of the document, where each sentence is separated and visible. All miRNA relations, extracted using relation extraction tools, are displayed at the bottom of the abstract. Each relation is also accompanied by the sentence number from which the relation was extracted.

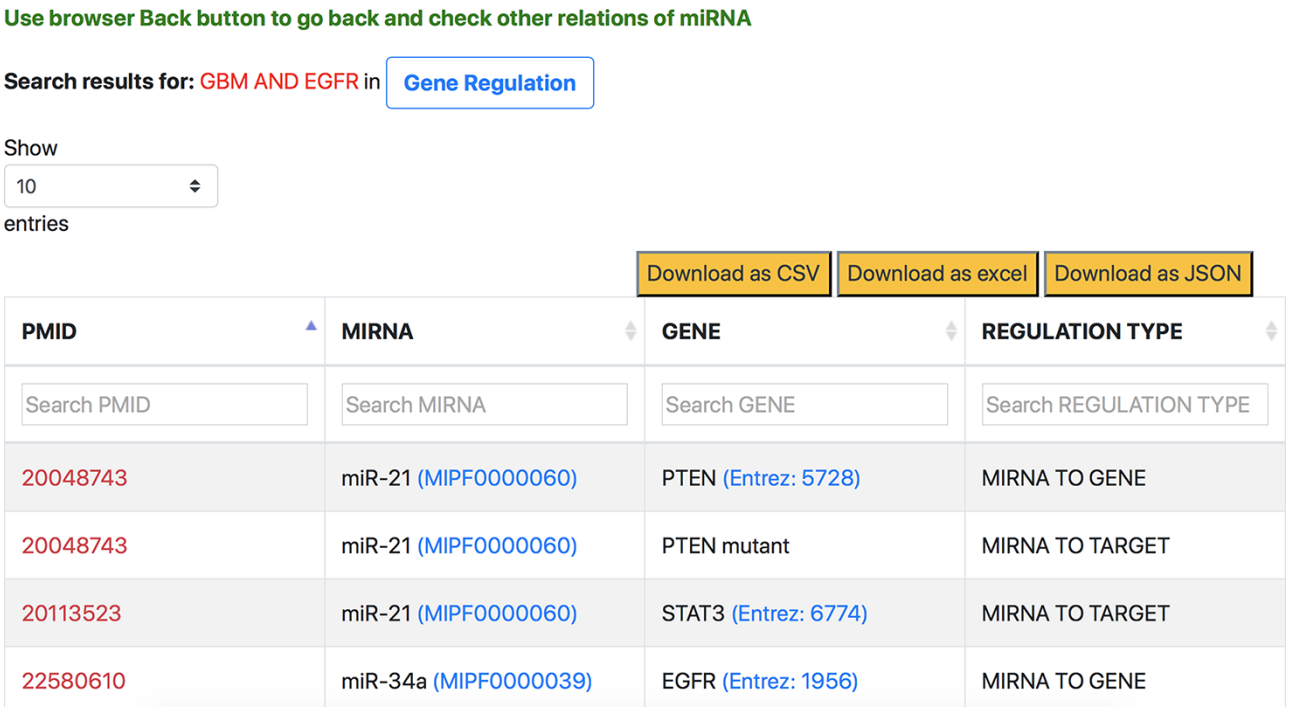


Figure 7. Response page of 'Gene Regulation' tab containing gene regulation information of miRNAs for context-centric query 'GBM AND EGFR'.

Additional features

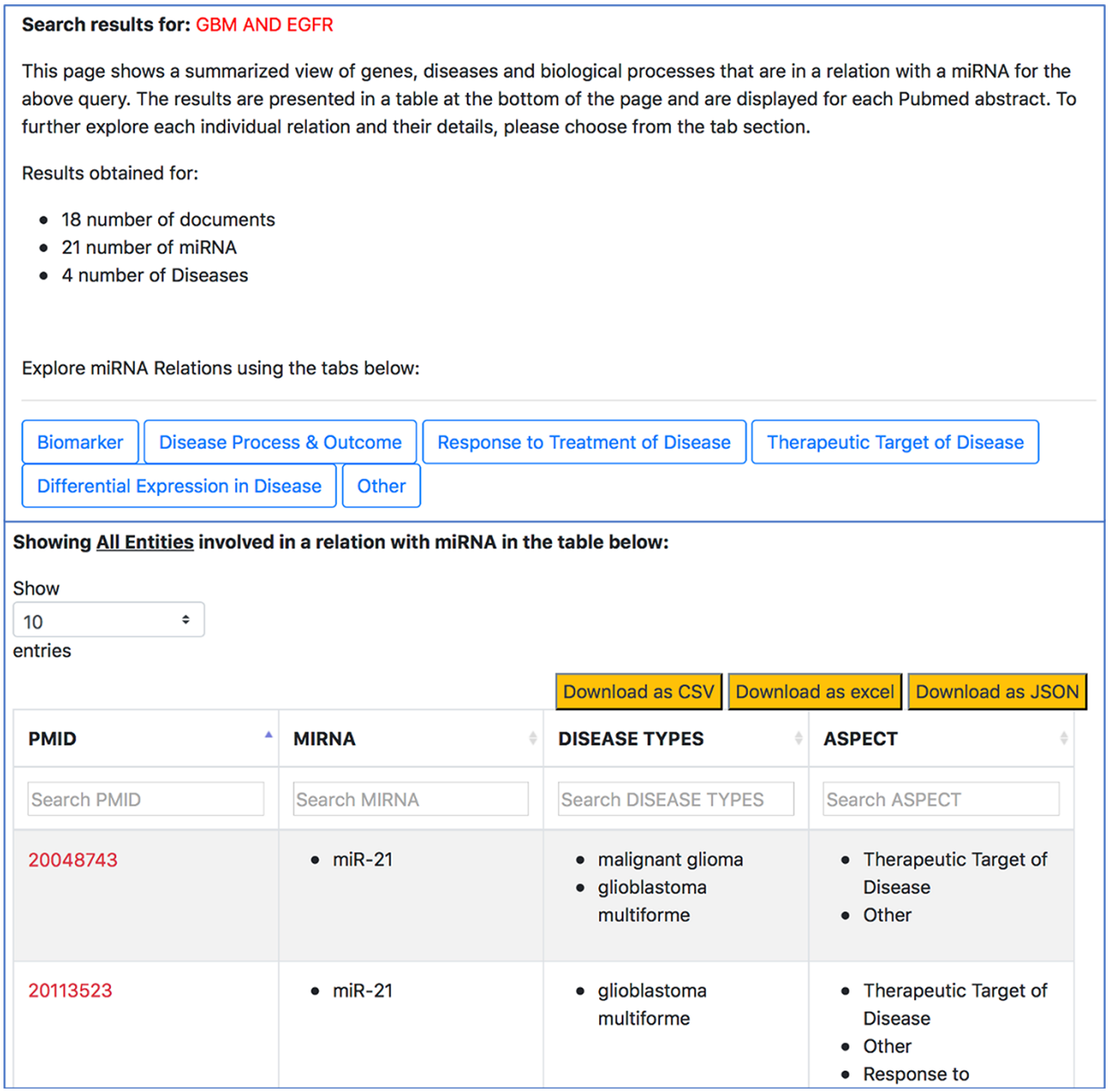
Sorting and filtering capabilities The tables in the high-level view and aspect-specific view can be sorted and filtered based on the user's information requirement (refer to Figure 10 in Section for Case Study 1 in Results and Discussion). Sorting on the column is performed by clicking on the arrow next to the column header, while filtering is performed by using the search box below the column header. The case study 1 in Results and Discussion section shows the usefulness of sorting and filtering the tables.

Ontology-driven search and link-out capabilities We normalize the entities of different types using publicly available and standard ontologies/databases, specifically to (i) ensure querying, sorting and filtering capabilities in our interface do not miss synonym of terms and (ii) expand the information scope of this resource by integrating with external resources that provide descriptions each entity, such as the genomic context of genes or sequences of miRNAs, and expand querying capabilities in other manually curated resources to broaden the understanding of the user about the role of an miRNA. For example, in Figure 7, when a user looks at the specific miRNA–gene relation and they want to know more about the miRNA's sequence or the description of the gene, they can simply click on the entity term. A separate page on the browser takes them to an external ontology that describes the specific entities. Plus, normalizing the entities improves the filtered results from the tables. In Figure 10, the normalized gene terms ensure that different names of 'PTEN', such as 'phosphatase and tensin homolog' are also captured when the table is filtered.

Download functionality Users can download the results from the summarized table from the high-level view and relation-specific tables from the aspect-centric view as a JSON file, CSV file or an Excel file.

Table 2. Summarized information of the number of bioentities and number of connections between miRNA and other bioentities in emiRIT

Entity and entity pairs in a relation		Number of instances
miRNAs		3099
Genes		15,486
Biological processes		1300
Diseases (cancer)		255
Extracellular locations (transporters and biofluids)		52
miRNA gene	Total	69,152
	miRNA target	34,302
	Human miRNA target	29,608
miRNA–biological processes		23,390
miRNA disease		12,300
miRNA extracellular locations		3782



Search results for: GBM AND EGFR

This page shows a summarized view of genes, diseases and biological processes that are in a relation with a miRNA for the above query. The results are presented in a table at the bottom of the page and are displayed for each Pubmed abstract. To further explore each individual relation and their details, please choose from the tab section.

Results obtained for:

- 18 number of documents
- 21 number of miRNA
- 4 number of Diseases

Explore miRNA Relations using the tabs below:

Biomarker | Disease Process & Outcome | Response to Treatment of Disease | Therapeutic Target of Disease | Differential Expression in Disease | Other

Showing All Entities involved in a relation with miRNA in the table below:

Show 10 entries

Download as CSV | Download as excel | Download as JSON

PMID	MIRNA	DISEASE TYPES	ASPECT
Search PMID	Search MIRNA	Search DISEASE TYPES	Search ASPECT
20048743	• miR-21	• malignant glioma • glioblastoma multiforme	• Therapeutic Target of Disease • Other
20113523	• miR-21	• glioblastoma multiforme	• Therapeutic Target of Disease • Other • Response to

Figure 8. Disease-oriented information context-centric query 'GBM AND EGFR'.

Results and discussion

We collected around 121 371 miRNA-related abstracts from PubMed, out of which we extracted more than 149 233 miRNA–PMID pairs. 49 010 out of 121 371 abstracts contained relations between miRNAs and other bioentities. Table 2 shows the number of unique entities in these 49 010 abstracts as well as the number of relations between miRNAs and other entities.

While the number of connections express how much emiRIT extracted from the literature, it does not give us a sense of what information has been missed in abstracts. Therefore, we decided to conduct case studies to assess the information coverage of different aspects of miRNAs through our resource. Since emiRIT disease captures only cancer, we use review articles for our case studies to find that the extent of information emiRIT is able to capture. We assume that the users of our resource will look for miRNA information in some context. As miRNAs are mostly explored in the context of diseases, we decided to focus our case studies on two widely investigated aspects of miRNAs that are important to understand miRNA's role in diseases—the target information and the differential expression of miRNAs. We use two highly cited review articles in these case studies, where one of them explores

PMID: 17297439

Title : MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells

Abstract :

1. Neuroblastoma (NB) is one of the most common forms of cancer in children, accounting for 15% of pediatric cancer deaths
2. The clinical course of these tumors is highly variable and is dependent on such factors as age at presentation, stage, ploidy and genomic abnormalities
3. Hemizygous deletion of chromosome 1p occurs in approximately 30% of advanced stage tumors, is associated with a poor prognosis, and likely leads to the loss of one or more tumor suppressor genes
4. We show here that microRNA (miRNA)-34a (1p36.23) is generally expressed at lower levels in unfavorable primary NB tumors and cell lines relative to normal adrenal tissue and that reintroduction of this miRNA into three different NB cell lines causes a dramatic reduction in cell proliferation through the induction of a caspase-dependent apoptotic pathway
5. As a potential mechanistic explanation for this observation, we demonstrate that miR-34a directly targets the messenger ribonucleic acid (mRNA) encoding E2F3 and significantly reduces the levels of E2F3 protein, a potent transcriptional inducer of cell-cycle progression
6. Furthermore, miR-34a expression increases during retinoic acid-induced differentiation of the SK-N-BE cell line, whereas E2F3 protein levels decrease
7. Thus, adding to the increasing role of miRNAs in cancer, miR-34a may act as a suppressor of NB tumorgenesis

Gene Regulation

MIRNA	GENE	REGULATION TYPE	SENTENCE
miR-34a (MIPF0000039)	E2F3 (Entrez: 1871)	MIRNA TO TARGET	5

Involvement in Biological Process

MIRNA	BIOLOGICAL PROCESS	PATHWAY	SENTENCE
miR-34a (MIPF0000039)	positive regulation of apoptotic process (GO:0043065)		0

Disease: Other

MIRNA	DISEASE TYPES	SENTENCE
miR-34a (MIPF0000009)	neuroblastoma (DOID:769)	7

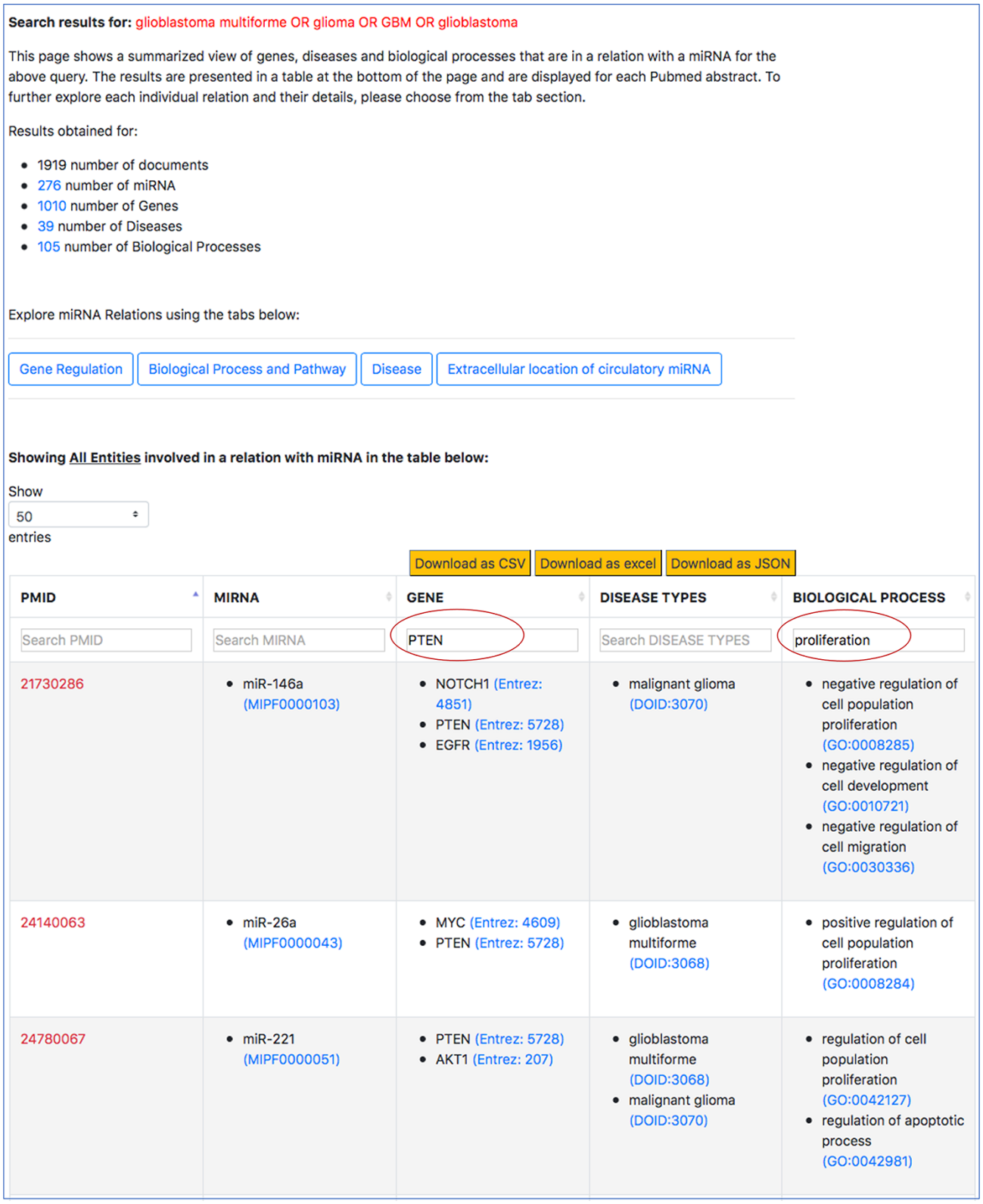
Figure 9. Document-specific view for PMID 17 297 439.

miRNA's role in the context of glioblastoma multiforme (GBM) and the other explores differential expression in human gastric cancer. In the absence of highly cited review articles on the other miRNA aspects in the context of cancer, we limit our case studies to the above-mentioned aspects. However, we also conduct a further case study to assess the usage of emiRIT in the curation of miRNA information.

Case study 1: target information of miRNAs in the context of a disease

The first case study explores the relation of miRNAs with bioentities in the context of GBM. A recent comprehensive review explored miRNA's role in eight hallmarks of GBM (62). We explored the first two hallmarks of GBM in our case study and compared the findings using our interface with that provided in the review.

The first hallmark question in the review investigates the aberrant miRNAs that affect receptor tyrosine kinase (RTK) signaling networks to promote cell proliferation in GBM cells. The review identified 29 miRNA–target pairs described as the miRNAs regulating the targets and modulating the RTK signaling network leading to cell proliferation. We captured 25 of the 29 miRNA–target pairs mentioned in the review using our interface. We used the general keyword search with the query (glioblastoma multiforme OR glioma OR GBM OR glioblastoma).



Search results for: glioblastoma multiforme OR glioma OR GBM OR glioblastoma

This page shows a summarized view of genes, diseases and biological processes that are in a relation with a miRNA for the above query. The results are presented in a table at the bottom of the page and are displayed for each Pubmed abstract. To further explore each individual relation and their details, please choose from the tab section.

Results obtained for:

- 1919 number of documents
- 276 number of miRNA
- 1010 number of Genes
- 39 number of Diseases
- 105 number of Biological Processes

Explore miRNA Relations using the tabs below:

Gene Regulation | Biological Process and Pathway | Disease | Extracellular location of circulatory miRNA

Showing All Entities involved in a relation with miRNA in the table below:

Show 50 entries

Download as CSV | Download as excel | Download as JSON

PMID	MIRNA	GENE	DISEASE TYPES	BIOLOGICAL PROCESS
Search PMID	Search MIRNA	PTEN	Search DISEASE TYPES	proliferation
21730286	• miR-146a (MIPF0000103)	• NOTCH1 (Entrez: 4851) • PTEN (Entrez: 5728) • EGFR (Entrez: 1956)	• malignant glioma (DOID:3070)	• negative regulation of cell population proliferation (GO:0008285) • negative regulation of cell development (GO:0010721) • negative regulation of cell migration (GO:0030336)
24140063	• miR-26a (MIPF0000043)	• MYC (Entrez: 4609) • PTEN (Entrez: 5728)	• glioblastoma multiforme (DOID:3068)	• positive regulation of cell population proliferation (GO:0008284)
24780067	• miR-221 (MIPF0000051)	• PTEN (Entrez: 5728) • AKT1 (Entrez: 207)	• glioblastoma multiforme (DOID:3068) • malignant glioma (DOID:3070)	• regulation of cell population proliferation (GO:0042127) • regulation of apoptotic process (GO:0042981)

Figure 10. Filtered search result for 'PTEN' in the context of cell proliferation in GBM.

Following is an example of our mode of search vthat we conducted on the interface to retrieve 25 out of the 29 miRNA–target pairs. For this example, we will consider 'PTEN', which was found to be a target for increased tumor growth by the review of oncomiRs, such as miR-17-5p, miR-19a/b, miR-21, miR-1908, miR-494-3p,

miR-10a/10b, miR-23a and miR-26a. To observe how many of these miRNA-'PTEN' pairs were also retrieved using emiRIT, we filtered the aforementioned general query search result in the interface with 'PTEN' as the keyword. We then observed all the miRNAs for any abstract that mentioned 'proliferation'. The snapshot of a filtered table using 'PTEN' in the search box above the table is shown in Figure 10.

We were able to identify miR-19a, miR-26a, miR-494-3p and miR-21 directly from the rows that contained both 'PTEN' and 'proliferation'. miR-23a was identified in rows that contained 'PTEN', while miR-10 was found in the same row as 'PTEN' and 'migration'. miR-17-5p was found to target PTEN but no mention of 'proliferation' or terms associated with 'proliferation' were detected by our resource. Even though we were able to find miR-1908 in the context of GBM and proliferation, we could not find PTEN to be the target of miR-1908. Additionally, we were able to capture much more information than what was found in the review. In the last two rows shown in Figure 10, we not only see that miR-26a was found to be in a relation with PTEN and was involved in proliferation, we also find other miRNAs, for example, miR-221, that were associated with PTEN and proliferation. The review did not identify miR-221 and PTEN as an miRNA–target pairs involved in cell proliferation. The abstract evidence, shown in Figure 11, suggests that miR-221 targets PTEN in the context of cell proliferation (sentences 6–10) in GBM.

Case study 2: differential expression of miRNAs in the context of a disease

Our second case study explores the differential miRNA expression in human gastric cancer. We used a highly cited review by Shrestha *et al.*, 2014 (63), which surveys miRNA expression profiling studies in human gastric cancer. The review mentions 41 miRNAs that were consistently upregulated and 28 miRNAs that were consistently downregulated. For our case study, we looked at these 69 miRNAs to see how many of them are found to be upregulated or downregulated using emiRIT.

On our interface, we used the disease-centric search with the query 'stomach cancer' from the list of disease ontology diseases autocompleted by the NCBO BioPortal widget. We use 'stomach cancer' because a search for 'gastric cancer' in the disease ontology also provided results for 'stomach cancer'. We then navigated to the 'Differential Expression in Disease' tab from the 'Disease' tab. Figure 12 shows a screenshot of a table in the resulting page from the 'Differential Expression in Disease' tab. We filtered the information on this table by using the second column for miRNAs and searched for the 69 miRNAs identified in the review. Supplementary Tables S3 and S4 compare between our findings and that of the review for all 69 miRNAs.

Observations on 41 upregulated and 28 downregulated miRNAs

To compare our findings with that of the review, we manually explored the papers that the review cited. We found

PMID: 24780067

Title : MicroRNA-221 targeting PI3-K/Akt signaling axis induces cell proliferation and BCNU resistance in human glioblastoma

Abstract :

1. MicroRNAs (miRNAs) are short regulatory RNAs that negatively regulate protein biosynthesis at the post-transcriptional level and participate in the pathogenesis of different types of human cancers, including glioblastoma
2. In particular, the levels of miRNA-221 are overexpressed in many cancers and miRNA-221 exerts its functions as an oncogene
3. Nevertheless, the roles of miRNA-221 in carmustine (BCNU)-resistant glioma cells have not been totally elucidated
4. In the present study, we explored the effects of miRNA-221 on BCNU-resistant glioma cells and the possible molecular mechanisms by which miRNA-221 mediated the cell proliferation, survival, apoptosis and BCNU resistance were investigated
5. We found that miR-221 was overexpressed in glioma cells, including BCNU-resistant cells
6. Moreover, we found that miR-221 regulated cell proliferation and BCNU resistance in glioma cells
7. Overexpression of miR-221 led to cell survival and BCNU resistance and reduced cell apoptosis induced by BCNU, whereas knockdown of miR-221 inhibited cell proliferation and prompted BCNU sensitivity and cell apoptosis
8. Further investigation revealed that miR-221 down-regulated PTEN and activated Akt, which resulted in cell survival and BCNU resistance
9. Overexpression of PTEN lacking 3'UTR or PI3-K/Akt specific inhibitor wortmannin attenuated miR-221-mediated BCNU resistance and prompted cell apoptosis
10. We propose that miR-221 regulated cell proliferation and BCNU resistance in glioma cells by targeting PI3-K/PTEN/Akt signaling axis
11. Our findings may provide a new potential therapeutic target for treatment of glioblastoma

Figure 11. Abstract evidence for 'miR-221' promoting 'proliferation' by targeting 'PTEN' in the context of GBM.

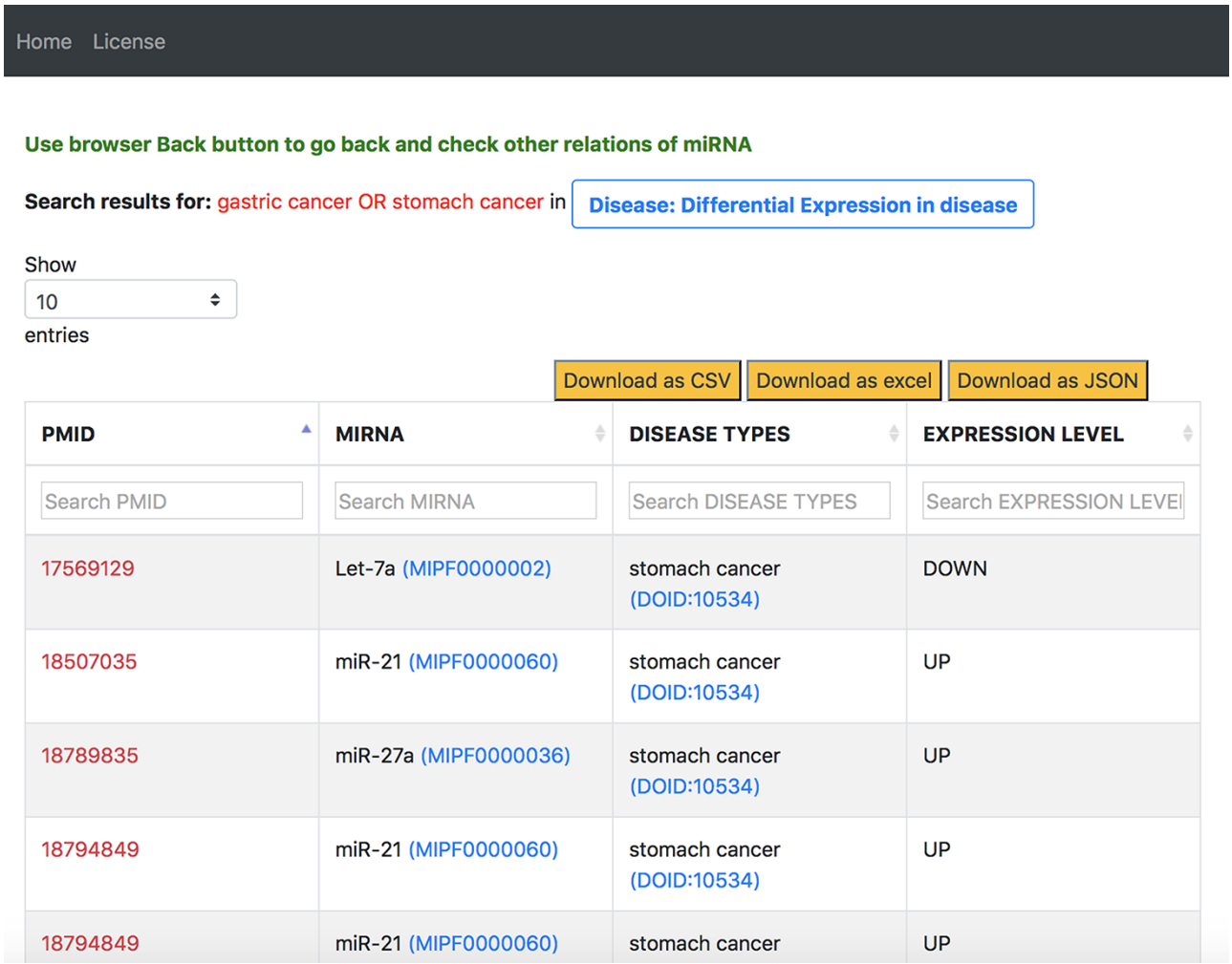


Figure 12. Output response page of 'Differential Expression in Disease' tab for 'gastric cancer OR stomach cancer' query in general keyword-centric search mode.

that all the miRNAs, except one, were mentioned in tables, supplementary tables, figures and the whole text beyond the abstracts of papers cited in the review. Since the miRNAs often occur in supplementary tables in these papers and there is no guarantee that the miRNA will occur in the text of the paper, we decided to go for alternate sources of gastric cancer-related papers to find the most consistent upregulated and downregulated miRNAs in the context of gastric cancer.

Observations on 41 upregulated miRNAs

Observation 1: Consistent with review Out of the 41 upregulated miRNAs, we found 28 miRNAs were upregulated in gastric cancer and stomach cancer.

Observation 2: Inconsistent with review

Downregulated miRNA sequences. Contrary to the review's description, we found 7 out of the 41 miRNAs were consistently downregulated in multiple abstracts. We further analyzed these abstracts and we found that the miRNAs were indeed downregulated. For example, miR-7 was found to be downregulated in PMIDs 22614005 (64), 24573489 (65), 26798443 (66). Similarly, miR-200b was found downregulated in PMIDs 23995857 (67) and 30999814 (68).

Closely related upregulated miRNA sequences From the remaining six miRNAs, we found that closely related miRNA sequences for five of them were upregulated. The review mentioned miR-18b was upregulated (Table S2 in supplementary tables). However, we found evidence for miR-18a to be upregulated and not for miR-18b. Similarly, we found miR-199a to be upregulated instead of miR-199-5p, miR-301a to be upregulated instead of miR-301, Let-7d and Let-7f to be upregulated instead of Let-7i. We found miR-519 in place of miR-519d in the context of stomach cancer but the expression level was downregulated.

However, miR-1259 was not detected by our tools. Since the review mostly finds upregulated miRNAs from tables or supplementary tables of the papers they survey and our current relation extraction tools are limited to abstracts, we are only able to extract upregulated miRNA mentions from abstracts.

Observations on 28 downregulated miRNAs

Observation 1: Consistent with review We found 20 out of 28 downregulated miRNAs mentioned in the review using emiRIT.

Observation 2: Inconsistent with review

Upregulated miRNA sequences. From the remaining eight miRNAs, miR-150 and miR-139 were found to be upregulated.

Closely related downregulated miRNA sequences Instead of miR-320 c mentioned in the review, we found miR-320 to be downregulated. Similarly, miR-30b was found to be downregulated instead of miR-30d. However, miR-30d was found to be downregulated in the context of large-intestine cancer and colorectal cancer.

We could not find the remaining four miRNAs in the context of stomach cancer or gastric cancer from abstracts of miRNA papers.

Observations on the top 3 upregulated and downregulated miRNAs in the review

The review also stated that from the list of upregulated miRNAs, the most consistently reported miRNAs were miR-21, followed by miR-25, miR-92 and miR-223. From the list of downregulated miRNAs, the authors found that the most consistently reported miRNAs were miR-375, miR-148a followed by miR-638. To find the most consistent upregulated and downregulated miRNAs using emiRIT, we conducted an additional search on the table shown in Figure 12. We filtered this table further to consider only the upregulated or only the downregulated cases and downloaded the resultant tables as excel separate files. In each file, then looked at the total number of abstracts supporting the regulation (up or down) of an miRNA.

Consistent with the review's findings, we found miR-21 to be the most reported miRNA upregulated in stomach cancer, extracted from 22 abstracts in emiRIT. The next most frequently reported miRNAs upregulated in abstracts using emiRIT were miR-25 (seven abstracts), miR-27a (seven abstracts), miR-223 (six abstracts), miR-20a (six abstracts) and miR-214 (six abstracts). While miR-92 was not in the top list, we did identify it in three abstracts. From the downregulated miRNAs, again consistent with the review's findings, we found miR-375 to be the most frequently reported miRNA downregulated in stomach cancer, occurring in nine abstracts in our resource, followed by miR-148a (8 abstracts), miR-218 (6 abstracts) and miR-133b (6 abstracts).

Case study 3: usage of emiRIT in miRNA information curation

We wanted to assess the usage of emiRIT in the curation of miRNA information. For this purpose, we use the GO annotations data set curated by Huntley *et al.*, 2018 (69). The dataset was downloaded from https://www.ebi.ac.uk/QuickGO/annotations?extension=has_input&geneProductType=miRNA&assignedBy=ARUK-UCL,BHF-UCL,ParkinsonsUK-UCL. We observed that out of 369 PMIDs curated by the UCL group, emiRIT extracted miRNA information from 368 PMIDs, irrespective of whether a disease was mentioned or not in the abstracts.

The UCL group curates the publications to assign GO terms to miRNAs. To do a proper job of comparing their curation with information extracted using emiRIT, access to the full-length articles, experimental methods and expert assessment are needed. emiRIT does not attempt to do GO term annotations. To help users navigate the miRNA information in the literature, emiRIT just captures terms about cellular processes and verifies if the terms correspond to any GO term using exact string matching. For example, the UCL annotations for miR-378a-3p included involvement in the negative regulation of cytokine production involved in inflammatory response (GO:1900016) with evidence from PMID:31824476 (70). From this abstract, emiRIT extracts the mention of IL-33 being a target of miR-378. An expert with prior background knowledge might infer the GO term annotation based on the properties of this particular gene. However, emiRIT does not currently utilize any background knowledge and does not attempt to make any inferences. As stated before, emiRIT only captures terms about cellular processes and sees if it exactly matches a GO term. Perhaps, more information can be found in the full-length article. But the abstract of this paper does not mention 'cytokine production'.

This situation was replicated in 15 annotations of GO terms that were not about gene regulation. Thus, a full comparison of UCL GO annotations with emiRIT would not be appropriate. However, we noticed that abstracts are indeed good sources of miRNA–gene relations, and hence, we decided to include a comparison of our miRNA–gene extraction with the subset of UCL annotations that contained information about GO terms corresponding to gene regulation.

We looked at the GO annotations gene silencing by miRNA (GO:0035195), miRNA-mediated inhibition of translation (GO:0035278), gene silencing by RNA (GO:0031047) and negative regulation of gene expression (GO:0010629) with evidence code IDA. Additionally, we mapped the Uniprot Ids of the annotations to their corresponding NCBI Gene Ids from http://uniprot.org/, since emiRIT normalizes gene mentions to the NCBI Gene Id. We observed that for 261 annotations, the miRNA and gene mentions occurred in the same sentence and for the remaining 237 annotations, the miRNA and gene pairs either did not occur/co-occur in the same sentence in a particular abstract or did not occur anywhere in the abstract. Since emiRIT does not attempt to make inferences and uses textual patterns to extract relations between entity pairs in every sentence, we further analyzed the UCL 261 annotations and found that emiRIT correctly extracted 213 of them.

We next restricted our observations of the miRNA–gene pairs to a specific gene—PPARG and downloaded miRNA-PPARG annotations using https://www.ebi.ac.uk/QuickGO/annotations?extension=has_input(UniProtKB:P37231&geneProductType=miRNA. From this dataset, we found 15 papers that had an miRNA-PPARG annotation. We separately conducted a keyword search on our system using 'PPAR gamma' and observed that emiRIT could extract the miRNA-PPARgamma relations from 11 out of 15 PMIDs. In one of the remaining PMIDs, we observed that the abstract of PMID:20693317 (71) mentioned PPAR alpha instead of PPAR gamma and emiRIT was able to extract the relation between the miRNA and PPAR alpha.

The above case studies showed the amount of information that can be found from the text, specifically abstracts, using emiRIT. The advantage of using our resource is that users get a more comprehensive picture of the miRNA information for their specific requirements.

Conclusion and future work

In this paper, we have described emiRIT, a text-mined-based resource for miRNA information. We used different existing and in-house developed text-mining tools to capture connections between miRNA gene, miRNA disease (cancer), miRNA-biological process and pathways, and miRNA-extracellular locations. Furthermore, instead of just stating an association between an miRNA and a disease, we present a more detailed role of an miRNA in the context of disease by distinguishing between (i) impact of miRNA on disease process and outcome, (ii) influence of miRNA on disease treatment, (iii) diagnostic role of miRNAs as biomarkers, (iv) role of miRNAs as therapeutic targets in diseases, (v) others, where the particular role is not clear, but the miRNA is associated with a disease or regulates a disease.

The output of the different text-mining tools is combined, and the output format is unified for easy visualization and navigation of information. All the different miRNA connections are presented to the users via an interface at https://research.bioinformatics.udel.edu/emirit/. Here, the users can easily transition between these connections to get a broader understanding of the role of miRNAs and examine these roles in a context specific to their different information needs. Literature evidence is provided for every result at the abstract and sentence level, which not only increases the confidence of the results extracted from the text but allows users to explore the papers for additional background and experimental context of the results. miRNAs and other bioentities are normalized using standard ontologies to expand querying abilities and integrate with external resources.

We have conducted two case studies to show the extent of miRNA information that can be found through emiRIT. Since the primary function of miRNAs is to regulate gene expression and their dysregulation often leads to diseases, we focused on the target information and differential expressions of miRNAs in the context of diseases for our case studies. Since emiRIT disease currently includes only cancer, we relied on review articles for the case studies to find the extent of information emiRIT is able to capture. We conducted an additional case study to assess the usage of emiRIT in the curation of miRNA information.

In this paper, we attempt to provide an up-to-date and user-friendly resource to facilitate access to comprehensive miRNA information from the literature on a large scale, enabling users to exploit, interpret and connect existing knowledge to design new investigations and theories. In the future, we plan to extend to diseases other than cancer and to improve the relation extraction tool capturing the connections between miRNAs and extracellular locations. We also plan to integrate additional miRNA-entity relation knowledge from external databases and provide network visualization through Cytoscape. Finally, we plan to extend our resource to full-length PMC open access publications.

Supplementary data

Supplementary data are available at *Database* online.

References

1. Vidigal,J.A. and Ventura,A. (2015) The biological functions of miRNAs: lessons from in vivo studies. *Trends Cell Biol.*, **25**, 137–147.

2. Ardekani,A.M. and Naeini,M.M. (2010) The role of microRNAs in human diseases. *Avicenna J. Med. Biotechnol.*, **2**, 161.
3. Sun,W., Julie Li,Y.S., Huang,H.D. *et al.* (2010) microRNA: a master regulator of cellular processes for bioengineering systems. *Annu. Rev. Biomed. Eng.*, **12**, 1–27.
4. Sonntag,K.C. (2010) microRNAs and deregulated gene expression networks in neurodegeneration. *Brain Res.*, **1338**, 48–57.
5. Feng,Y.H. and Tsao,C.J. (2016) Emerging role of microRNA-21 in cancer. *Biomed. Rep.*, **5**, 395–402.
6. Ha,T.Y. (2011) microRNAs in human diseases: from cancer to cardiovascular disease. *Immune Netw.*, **11**, 135–154.
7. Ha,T.Y. (2011) microRNAs in human diseases: from autoimmune diseases to skin, psychiatric and neurodegenerative diseases. *Immune Netw.*, **11**, 227–244.
8. Hwang,H.W. and Mendell,J.T. (2006) microRNAs in cell proliferation, cell death, and tumorigenesis. *Br. J. Cancer*, **94**, 776.
9. Galperin,M.Y., Fernández-Suárez,X.M. and Rigden,D.J. (2017) The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. *Nucleic Acids Res.*, **45**, D1–D11.
10. Moore,A.C., Winkjer,J.S. and Tseng,T.T. (2015) Bioinformatics resources for microRNA discovery. *Biomark Insights*, **10**, BMI–S29513.
11. Canese,K., and Weis,S. (2013) PubMed: The bibliographic database. *The NCBI handbook [internet]*, 2nd edn. National Center for Biotechnology Information, US. https://www.ncbi.nlm.nih.gov/pubmed/.
12. Schetter,A.J., Leung,S.Y., Sohn,J.J. *et al.* (2008) microRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*, **299**, 425–436.
13. Ke,G., Liang,L., Yang,J.M. *et al.* (2013) MiR-181a confers resistance of cervical cancer to radiation therapy through targeting the pro-apoptotic PRKCD gene. *Oncogene*, **32**, 3019–3027.
14. Childs,G., Fazzari,M., Kung,G. *et al.* (2009) Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma. *Am. J. Pathol.*, **174**, 736–745.
15. Qin,W., Shi,Y., Zhao,B. *et al.* (2010) miR-24 regulates apoptosis by targeting the open reading frame (ORF) region of FAF1 in cancer cells. *PLoS One*, **5**, e9429.
16. Slaby,O., Svoboda,M., Fabian,P. *et al.* (2007) Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology*, **72**, 397–402.
17. Kutay,H., Bai,S., Datta,J. *et al.* (2006) Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. *J. Cell. Biochem.*, **99**, 671–678.
18. Orchard,S., Ammari,M., Aranda,B. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, 358.
19. Huntley,R.P., Kramarz,B., Sawford,T. *et al.* (2018) Expanding the horizons of microRNA bioinformatics. *RNA*, **24**, 1005–1017.
20. Karagkouni,D., Paraskevopoulou,M.D., Chatzopoulos,S. *et al.* (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.*, **46**, D239–D245.
21. Sticht,C., de la Torre,C., Parveen,A. *et al.* (2018) miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*, **13**, e0206239.
22. Chou,C.H., Shrestha,S., Yang,C.D. *et al.* (2017) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.
23. Xiao,F., Zuo,Z., Cai,G. *et al.* (2009) miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
24. Tong,Z., Cui,Q., Wang,J. *et al.* (2019) TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **47**, D253–D258.
25. Jiang,Q., Wang,Y., Hao,Y. *et al.* (2008) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
26. Wang,D., Gu,J., Wang,T. *et al.* (2014) OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics*, **30**, 2237–2238.
27. Huang,Z., Shi,J., Gao,Y. *et al.* (2018) HMDD v3.0: a database for experimentally supported human microRNA–disease associations. *Nucleic Acids Res.*, **47**, D1013–D1017.
28. Das,S.S., Saha,P. and Chakravorty,N. (2018) miRwayDB: a database for experimentally validated microRNA-pathway associations in pathophysiological conditions. *Database*, **2018**. 10.1093/database/bay023.
29. Gene Ontology Consortium. (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
30. Russo,F., Di Bella,S., Nigita,G. *et al.* (2012) miRandola: extracellular circulating microRNAs database. *PLoS One*, **7**, e47786.
31. Yang,Z., Wu,L., Wang,A. *et al.* (2017) dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.*, **45**, D812–D818.
32. Ruepp,A., Kowarsch,A., Schmidl,D. *et al.* (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.*, **11**, R6.
33. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
34. Szcześniak,M.W. and Makałowska,I. (2014) miRNEST 2.0: a database of plant and animal microRNAs. *Nucleic Acids Res.*, **42**, D74–D77.
35. Naeem,H., Küffner,R., Csaba,G. *et al.* (2010) miRSel: automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinform.*, **11**, 135.
36. Li,G., Ross,K.E., Arighi,C.N. *et al.* (2015) miRTex: a text mining system for miRNA-gene relation extraction. *PLoS Comput. Biol.*, **11**, e1004391.

37. Xie,B., Ding,Q., Han,H. *et al.* (2013) miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics*, **29**, 638–644.
38. Gupta,S., Ross,K.E., Tudor,C.O. *et al.* (2016) miRiaD: a text mining tool for detecting associations of microRNAs with diseases. *J. Biomed. Semantics*, **7**, 1–15.
39. Murray,B.S., Choe,S.E., Woods,M. *et al.* (2010) An in silico analysis of microRNAs: mining the miRNAome. *Mol. Biosyst.*, **6**, 1853–1862.
40. Salhi,A., Essack,M., Alam,T. *et al.* (2017) DES-ncRNA: a knowledgebase for exploring information about human micro and long noncoding RNAs based on literature-mining. *RNA Biol.*, **14**, 963–971.
41. Manning,C., Surdeanu,M., Bauer,J. *et al.* (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60.
42. Wei,C.H., Kao,H.Y. and Lu,Z. (2013) PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res.*, **41**, W518–W522.
43. Kibbe,W.A., Arze,C., Felix,V. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
44. Petri,V., Jayaraman,P., Tutaj,M. *et al.* (2014) The pathway ontology - updates and applications. *J. Biomed. Semantics*, **5**, 7.
45. Gremse,M., Chang,A., Schomburg,I. *et al.* (2010) The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res.*, **39**, D507–D513.
46. Huang,D., Qiu,S., Ge,R. *et al.* (2015) miR-340 suppresses glioblastoma multiforme. *Oncotarget*, **6**, 9257–9270.
47. Huang,Y., Qi,Y., Du,J.Q. *et al.* (2014) microRNA-34a regulates cardiac fibrosis after myocardial infarction by targeting Smad4. *Expert Opin. Ther. Targets*, **18**, 1355–1365.
48. Zhu,H.Q., Li,Q., Dong,L.Y. *et al.* (2014) microRNA-29b promotes high-fat diet-stimulated endothelial permeability and apoptosis in apoE knock-out mice by down-regulating MT1 expression. *Int. J. Cardiol.*, **176**, 764–770.
49. Coulouarn,C., Factor,V.M., Andersen,J.B. *et al.* (2009) Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene*, **28**, 3526–3536.
50. Wang,P., Cao,J., Liu,S. *et al.* (2017) Upregulated microRNA-429 inhibits the migration of HCC cells by targeting TRAF6 through the NF-kappaB pathway. *Oncol. Rep.*, **37**, 2883–2890.
51. Gupta,S., Dingerdissen,H., Ross,K.E. *et al.* (2018) DEXTER: disease-expression relation extraction from text. *Database*, **2018**, bay045.
52. Gupta,S., Mahmood,A.A., Ross,K. *et al.* (2017) Identifying comparative structures in biomedical text. *BioNLP*, **2017**, 206–215.
53. Park,N.J., Zhou,H., Elashoff,D. *et al.* (2009) Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin. Cancer Res.*, **15**, 5473–5477.
54. Anfossi,S., Babayan,A., Pantel,K. *et al.* (2018) Clinical utility of circulating non-coding RNAs—an update. *Nat. Rev. Clin. Oncol.*, **15**, 541.
55. Peng,Y., Gupta,S., Wu,C. *et al.* (2015) An extended dependency graph for relation extraction in biomedical texts. *Proc. BioNLP*, **15**, 21–30.
56. Zhou,R., Zhang,Y., Du,G. *et al.* (2018) Down-regulated let-7b-5p represses glycolysis metabolism by targeting AURKB in asthenozoospermia. *Gene*, **663**, 83–87.
57. Trifunov,S., Natera-de Benito,D., Exposito Escudero,J.M. *et al.* (2020) Longitudinal study of three microRNAs in Duchenne muscular dystrophy and Becker muscular dystrophy. *Front. Neurol.*, **11**, 304.
58. Mahmood,A.A., Rao,S., McGarvey,P. *et al.* (2017) eGARD: extracting associations between genomic anomalies and drug responses from text. *PLoS One*, **12**, e0189663.
59. Ren,J., Li,G., Ross,K. *et al.* (2018) iTextMine: integrated text-mining system for large-scale knowledge extraction from the literature. *Database*, **2018**. 10.1093/database/bay128.
60. Bourhis,P., Reutter,J. L., Suárez,F., and Vrgoč,D. (2017) JSON: Data model, query languages and schema specification. In: *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. p. 123–135. https://www.json.org/json-en.html.
61. Chodorow,K. (2013) *MongoDB: The Definitive Guide: Powerful and Scalable Data Storage*. O'Reilly Media, Inc.
62. Yu,W., Liang,S. and Zhang,C. (2018) Aberrant miRNAs regulate the biological hallmarks of glioblastoma. *Neuromolecular Med.*, **20**, 452–474.
63. Shrestha,S., Hsu,S.D., Huang,W.Y. *et al.* (2014) A systematic review of microRNA expression profiling studies in human gastric cancer. *Cancer Med.*, **3**, 878–888.
64. Zhao,X., Dou,W., He,L. *et al.* (2013) microRNA-7 functions as an anti-metastatic microRNA in gastric cancer by targeting insulin-like growth factor-1 receptor. *Oncogene*, **32**, 1363–1372.
65. Xie,J., Chen,M., Zhou,J. *et al.* (2014) miR-7 inhibits the invasion and metastasis of gastric cancer cells by suppressing epidermal growth factor receptor expression. *Oncol. Rep.*, **31**, 1715–1722.
66. Chen,W.Q., Hu,L., Chen,G.X. *et al.* (2016) Role of microRNA-7 in digestive system malignancy. *World J. Gastrointest. Oncol.*, **8**, 121–127.
67. Tang,H., Deng,M., Tang,Y. *et al.* (2013) miR-200b and miR-200c as prognostic factors and mediators of gastric cancer cell progression. *Clin. Cancer Res: An Official J. Am. Assoc. Cancer Res.*, **19**, 5602–5612.
68. Zhang,F., Li,Y., Xu,W. *et al.* (2019) Long non-coding RNA ZFAS1 regulates the malignant progression of gastric cancer via the microRNA-200b-3p/Wnt1 axis. *Biosci. Biotechnol. Biochem.*, **83**, 1289–1299.
69. Huntley,R.P., Sitnikov,D., Orlic-Milacic,M. *et al.* (2016) Guidelines for the functional annotation of microRNAs using the Gene Ontology. *RNA*, **22**, 667–676.

70. Dubois-Camacho,K., Diaz-Jimenez,D., De La Fuente,M. *et al.* (2019) Inhibition of miR-378a-3p by inflammation enhances IL-33 levels: a novel mechanism of alarmin modulation in ulcerative colitis. *Front. Immunol.*, **10**, 2449.

71. Sarkar,J., Gou,D., Turaka,P. *et al.* (2010) microRNA-21 plays a role in hypoxia-mediated pulmonary artery smooth muscle cell proliferation and migration. *Am. J. Physiol. Lung. Cell. Mol. Physiol.*, **299**, L861–L871.