# The utilization of the OmniSearch semantic search tool to explore various microRNA regulation mechanisms in osteoarthritis

Jingshan Huang
Computer Science Department, School of Computing
University of South Alabama
Mobile, Alabama 36688, U.S.A.

Bi Liu
Department of Orthopedics, Shenzhen Luohu People's Hospital
Shenzhen, Guangdong 518001, China

Yang Liu
Department of Orthopedics, Shenzhen People's Hospital
The Second Clinical Medical College at Jinan University
Shenzhen, Guangdong 518109, China

Ji Chen†
Department of Orthopedics, Shenzhen People's Hospital
The Second Clinical Medical College at Jinan University
Shenzhen, Guangdong 518109, China

***Abstract*—Osteoarthritis (OA) is the most common joint disease worldwide, resulting in severe joint pain and significantly decreased life quality in the elderly population. Due to the lack of effective medicine that can reverse the degeneration of articular cartilage, ultimately all OA patients need to receive an artificial joint replacement surgery. Although the pathogenesis and progression of OA has attracted a lot research activities, most regulation mechanisms at genetic and genomics level have not yet been completely understood. In this paper, we introduce the utilization of OmniSearch, a semantic search software tool, to facilitate the exploration of important roles performed by numerous microRNA molecules on OA disease process. Our promising experimental results have indicated that the methodologies described in this paper are able to help effectively unraveling critical microRNA regulation mechanisms. Consequently, we can facilitate clinical investigators' efforts in the understanding of OA pathogenesis, and thus contributing to early diagnosis and effective treatment in OA in the future.**

***Keywords*—osteoarthritis, microRNA, target gene, bio-ontology, semantic integration and search.**

## I. Introduction

Osteoarthritis (OA) is the most common joint disease worldwide, resulting in severe joint pain and significantly decreased life quality in the elderly population. OA affects the entire joint structure, including articular cartilage, subchondral bone, ligaments, synovial membrane, joint capsule, and articular muscle, and ultimately leads to the degeneration of articular cartilage, joint deformity, and defuctionalization. More seriously, due to the lack of effective medicine that can reverse the degeneration of articular cartilage, ultimately all OA patients need to receive an artificial joint replacement surgery. Although the pathogenesis and progression of OA has attracted many research activities ([1–8] for example), most regulation mechanisms at genetic and genomics level have not yet been completely elucidated. Towards this end, we will need to significantly enhance our understanding of the genetic, genomic, and epigenetic foundation underlying OA disease process. Thus, it is necessary to develop more advanced methodologies that are capable of integrating biological and computational approaches in a seamless manner; only this way will it be possible for us to explore a more accurate representation of biological processes that regulate OA development and progression.

We report in this paper our efforts in effectively utilizing software analysis (based upon domain ontologies and semantic technologies), aiming to better investigate important roles performed by microRNA (miR) regulations in OA.

The rest of this paper is organized as follows. Section II summarizes state-of-the-art research in OA and semantic technologies; Section III describes our innovative methodologies; Section IV reports current findings from our method along with discussion; and finally, Section V concludes with important future work.

## II. Related Work

### A. Related work in miR regulations in OA

Although far from being completely understood, the genetic regulation mechanisms performed by various miRs in OA have attracted many research activities.

To identify miRs involved in OA, Iliopoulos et al. [1] tested the expression of 365 miRs in articular cartilage obtained from patients with OA undergoing knee replacement surgery and from normal individuals with no history of joint disease. They identified 16 miRs that are differentially expressed in OA compared to normal cartilage. Jones et al. [2] linked miRs and OA in a study conducted in 2009 that compared miR expression in human cartilage and bone removed during knee replacements with expression in post-mortem specimens from donors with no previous joint pain. In this study, out of expression profiling of 157 human miRs, a total of 17 miRs were identified in cartilage and 30 in bone showing differential expression of greater than 4-fold in disease tissue compared to normal tissue.

† Corresponding and Co-First author (Email: chenji710727@163.com)

A specific miR, mir-140, was shown to be expressed only in cartilaginous tissues of the developing zebrafish in 2005 [3]. In 2006, Tuddenham et al. [4] reported that mir-140 was specifically expressed in cartilage tissues of mouse embryos during long and flat bone development, and they detected that histone deacetylase 4 was down-regulated by this miR. In 2009, Miyaki et al. [5] compared gene expression profiling using miR microarrays and quantitative polymerase chain reaction in human articular chondrocytes and human mesenchymal stem cells (MSCs). They showed that mir-140 regulated cartilage development and homeostasis, and it could be reduced in OA cartilage and in response to IL-1$\beta$, resulting from a chondrocyte differentiation-related expression pattern — such reduction might contribute to abnormal gene expression in OA.

Matsukawa et al. [6] confirmed that hsa-miR-125b was expressed in both normal and OA chondrocytes, with significantly lower expression in OA chondrocytes than in normal chondrocytes. Furthermore, IL-1$\beta$-induced up-regulation of ADAMTS-4 was suppressed by over-expression of hsa-miR-125b in human OA chondrocytes. Their results indicated that this miR played an important role in regulating the expression of ADAMTS-4 in human chondrocytes, makinghsa-miR-125b a novel therapeutic target in OA.

It was discovered by Vonk et al. [7] that overexpression of hsa-miR-148a increased COL2A1 and decreased MMP13 and ADAMTS5 gene expression. Overexpression of hsa-miR-148a was able to inhibit hypertrophic differentiation and increase the production and deposition of type II collagen by OA chondrocytes, which was accompanied by an increased retention of proteoglycans. This research suggested that hsa-miR-148a may be a potential disease-modifying compound in OA, as it promotes hyaline cartilage production.

Zhang et al. [8] have found that hsa-miR-210 targeted 3'-UTR of death receptor 6 (DR6) to inhibit its expression. Both hsa-miR-210 mimic and DR6 siRNA transfection inhibited the activation of NF-KB pathway and cell apoptosis of chondrocytes. During the in-vivo study, an OA model was established on rats by anterior cruciate ligament transection (ACLT), and hsa-miR-210 expression was reduced in OA rats. In addition, hsa-miR-210 over-expressing lentivirus was injected into the OA rats as well. Their experimental results indicated that cytokines production, NF-KB, and DR6 expression in OA rats were all inhibited by hsa-miR-210 overexpression.

### B. Related work in semantic technologies

In biomedical investigation, when we need to integrate a large number of data sources that have heterogeneous semantics (different meanings), semantic technologies that are based on domain ontologies can render significant assistance.

Bio-ontologies have been widely utilized nowadays. Herein we briefly review some bio-ontologies that are related to our research in this paper. Gene Ontology (GO) [9] is probably the most successful and widely used bio-ontology, which contains three independent sub-ontologies: biological processes, molecular functions, and cellular components. Non-Coding RNA Ontology (NCRO) [10] [11] is an Open Biological and Biomedical Ontologies (OBO) [12] candidate reference ontology designed for non-coding RNA (ncRNA) domain. Ontology for MIcroRNA Target (OMIT) [13–15] serves as an application ontology to provide the community with common data elements and data exchange standards in the miR research.

An important part of our proposed methodologies in this paper is closely related to semantic search [16], which usually requires the utilization of structured knowledge to interpret search queries, through formal logic for example. One popular idea in semantic search systems ([17–21] for example) is to expand the query keywords utilizing synonyms and other relations not originally part of the query. A second way to implement semantic search [16] is to translate the original keyword-based search into some formal semantic queries through the adoption of domain ontologies.

## III. Methodologies

### A. Overview of our methodologies

- Step One. We will first conduct manual background literature search to identify a list of miRs that are likely involved in the OA disease process control.
- Step Two. For each miR returned from Step One, we will use the OmniSearch software tool [15] [22] to retrieve a set of its target genes (mRNAs). These putative target genes will be ranked using various criteria, which can be easily customized by users. Target genes will then be returned to users as computationally predicted results, which can further be biologically verified through wet-lab experiments. Note that biological experimental design is not included as part of the methodologies presented in this paper; rather, it is considered as our important future work, and Section V contains greater details in this regard.

### B. Domain ontologies underlying the OmniSearch system

OmniSearch is a typical semantic integration and search system, developed on NCRO and OMIT ontologies.

### C. OmniSearch software architecture

The overall software architecture in OmniSearch is demonstrated in Fig. 1, with the following working protocol:

- Query parameters are sent from the client's browser to the Apache server [23] through Ajax requests.
- SPARQL Protocol and RDF Query Language (SPARQL) [24] queries are dynamically generated by the Apache server using these query parameters, which are then sent to the Apache Jena Fuseki server.
- JSON objects, containing the requested information, are retrieved from the resource description framework (RDF) [25] triple store (installed on the Apache Jena Fuseki server) after running the dynamically generated SPARQL queries.
- These JSON objects are returned to the Apache server, which are used to generate either (1) a list of miRs and/or MeSH terms or (2) the HTML Markup for the search result table.
- Finally, the Apache server sends the obtained data, or an error message if the search fails, back to the

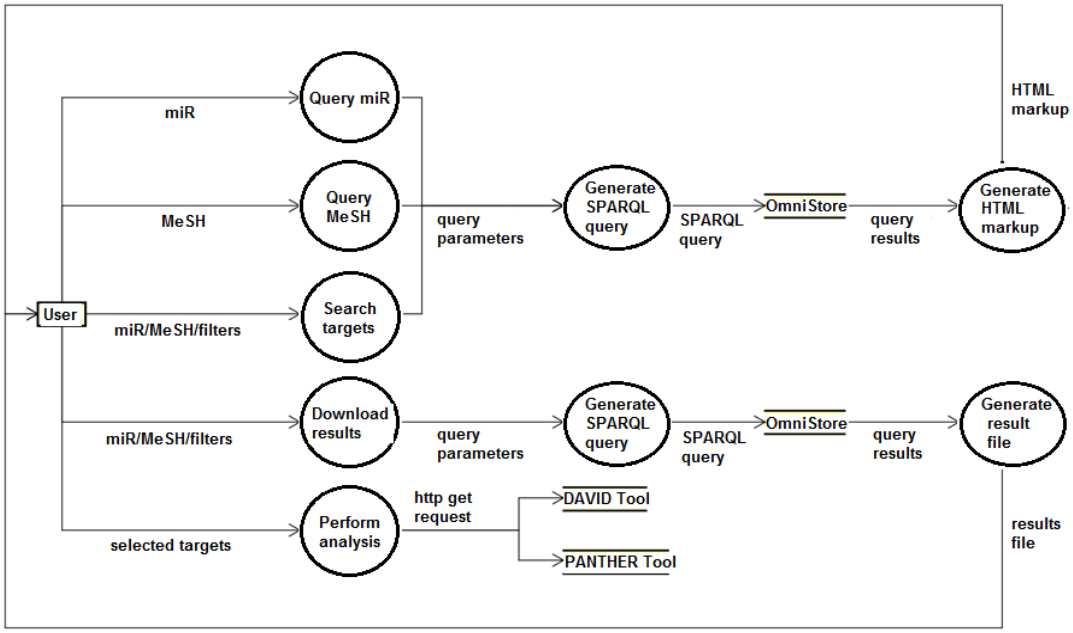


Fig. 1. Semantic search architecture in the OmniSearch system.

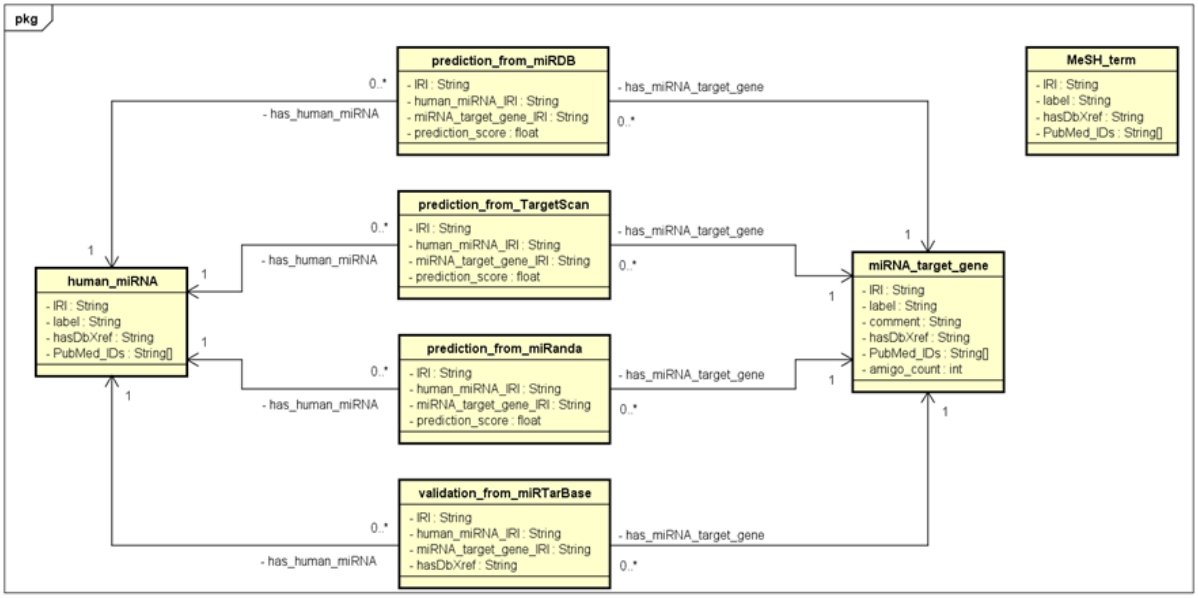


Fig. 2. Various data sources integrated in the OmniSearch system.

client's browser as a JSON object.

There are a total of ten data sources that were integrated in the OmniSearch system:

- Three miR target prediction databases: miRDB [26], TargetScan [27], and miRanda [28]. Each of these three databases contains computationally predicted target genes given a miR of interest.
- miRTarBase [29]: All target genes included in this database have already been biologically validated by some research groups.
- miRBase [30]: This is a database of published miR sequences and annotations. Each entry in the database represents a predicted hairpin portion of a miR transcript, along with the information on the location and sequence of the respective mature miR sequence.
- Three National Center for Biotechnology Information (NCBI) databases: PubMed [31], NCBI Gene [32], and MeSH [33]. These databases provide access to PubMed publications, various gene-related information, and MeSH term information, respectively.
- GO annotation database [34]: This database contains statements that describe the functions of specific genes

TABLE I. BACKGROUND LITERATURE SEARCHING RESULTS WITH REGARD TO CANDIDATE MIRS IN OA

| Putative Regulating miR | Target Gene | Signaling Pathway Involved |
|---|---|---|
| hsa-miR-140-5p | ADAMTS5 | IL-1$\beta$ |
| | MMP-13 | IGF |
| | MMP-13 | NF-B |
| | IGFBP-5 | TGF-$\beta$ |
| | DNPEP | BMP |
| | SMAD3 | TGF-$\beta$ |
| | RALA | SOX9 |
| | SP1 | BMP |
| hsa-miR-146a-5p | SMAD4 | TGF-$\beta$ |
| | IRAK1 | NF-B |
| | TRAF6 | NF-B |
| | MMP-13 | IL-1$\beta$ |
| hsa-miR-27a-3p | MMP-13 | NF-B |
| hsa-miR-34a-5p | SIRT1 | P53 |
| | COL2A1 | IL-1$\beta$ |
| hsa-miR-223-5p | PEX-16 | IL-1$\beta$ |
| hsa-miR-125b-5p | ADAMTS4 | IL-1$\beta$ |
| hsa-miR-675-5p | COL2A1 | TNF-$\alpha$ |
| hsa-miR-101-3p | ECM | SOX9 |
| hsa-miR-148a-5p | DNMT1 | methylation |
| hsa-miR-558 | COX-2 | NF-B |
| hsa-miR-455-5p | SMAD2/3 | TGF-$\beta$ |
| hsa-miR-29a-3p | SMAD2/3/4 | TGF-$\beta$ |
| | Axin2 | WNT |
| hsa-miR-145-5p | ECM | SOX9 |
| hsa-miR-149-5p | ADAMTS5 | TNF-$\alpha$ |
| hsa-miR-26a-5p | INOS | NF-B |
| hsa-miR-483-5p | MMP-13 | TGF-$\beta$ |
| hsa-miR-93-5p | OSM | AKT |
| hsa-miR-193b-5p | SOX9 | TGF-$\beta$ |
| hsa-miR-33a-3p | SREBP-2 | TGF-$\beta$ |
| hsa-miR-210-5p | DR6 | NF-B |
| hsa-miR-199a-5p | COX2 | P38/MAPK |
| hsa-miR-9-5p | MMP-13 | IL-1$\beta$ |
| hsa-miR-355-5p | COL10A1 | WNT/$\beta$-catenin |
| hsa-miR-22-5p | PPAR-$\alpha$ | IL-1$\beta$ |
| hsa-miR-24-3p | P16INK4A | IL-1$\beta$ |
| hsa-miR-21-5p | GDF-5 | TGF-$\beta$ |
| hsa-miR-98-5p | ADAMTS5 | IL-1$\beta$ |
| hsa-miR-127-3p | MMP13 | IL-1$\beta$ |
| hsa-miR-488-5p | ZIP8 | IL-1$\beta$ |
| hsa-miR-181b-5p | MMP13 | IL-1$\beta$ |

using GO concepts, and each statement is based on a specified piece of evidence.

- RNAcentral database [35]: A public resource that offers integrated access to a comprehensive and up-to-date set of non-coding RNA sequences.

Some of the above data sources as well as their interrelationship are demonstrated in Fig. 2.

## IV. RESULTS AND DISCUSSION

### A. Results from Step One — Searching background literature

Table I demonstrates the searching results on background literature, where a total of 30 candidate miRs were found to likely regulate OA disease process.

### B. Results from Step Two — Utilizing the OmniSearch tool

The OmniSearch software[1] is housed on a server with the following configuration: Intel(R) Core(TM) i7-3632 QM CPU @ 2.80 GHz 2.80 GHz; 32.00 GB memory; and Windows Server 8 Operating System. All results reported in this section were conducted on personal computers with the following configuration: Intel(R) Core(TM) i7-3632 QM CPU @ 2.50 GHz 2.50 GHz; 16.00 GB memory; and Windows 10 64-bit Operating System.

[1]OmniSearch is accessible at: http://omnisearch.soc.southalabama.edu/ui

TABLE II. DIFFERENT NUMBERS OF PUTATIVE TARGET GENES RETURNED BY MANUAL SEARCH AND OMNISEARCH

| Regulating miR | # of Putative Target Genes by Manual Search | # of Putative Target Genes by OmniSearch | % of Increased Target Gene Number |
|---|---|---|---|
| hsa-miR-140-5p | 8 | 25 | 213% |
| hsa-miR-146a-5p | 4 | 13 | 225% |
| hsa-miR-27a-3p | 1 | 2 | 100% |
| hsa-miR-34a-5p | 2 | 3 | 50% |
| hsa-miR-223-5p | 1 | 3 | 200% |
| hsa-miR-125b-5p | 1 | 3 | 200% |
| hsa-miR-675-5p | 1 | 2 | 100% |
| hsa-miR-101-3p | 1 | 3 | 200% |
| hsa-miR-148a-5p | 1 | 2 | 100% |
| hsa-miR-558 | 4 | 9 | 125% |
| hsa-miR-455-5p | 1 | 3 | 200% |
| hsa-miR-29a-3p | 2 | 5 | 150% |
| hsa-miR-145-5p | 1 | 3 | 200% |
| hsa-miR-149-5p | 1 | 3 | 200% |
| hsa-miR-26a-5p | 1 | 2 | 100% |
| hsa-miR-483-5p | 1 | 2 | 100% |
| hsa-miR-93-5p | 1 | 3 | 200% |
| hsa-miR-193b-5p | 1 | 2 | 100% |
| hsa-miR-33a-3p | 1 | 3 | 200% |
| hsa-miR-210-5p | 1 | 3 | 200% |
| hsa-miR-199a-5p | 1 | 5 | 400% |
| hsa-miR-9-5p | 1 | 2 | 100% |
| hsa-miR-355-5p | 1 | 3 | 200% |
| hsa-miR-22-5p | 1 | 3 | 200% |
| hsa-miR-24-3p | 1 | 2 | 100% |
| hsa-miR-21-5p | 1 | 2 | 100% |
| hsa-miR-98-5p | 1 | 2 | 100% |
| hsa-miR-127-3p | 1 | 3 | 200% |
| hsa-miR-488-5p | 1 | 2 | 100% |
| hsa-miR-181b-5p | 1 | 3 | 200% |

TABLE III. THE SYSTEM TIME AND SAVED USER TIME DURING MIR QUERYING

| Query | miR | System Time (seconds) | User Time (seconds) | Percentage of Saved Time for Users |
|---|---|---|---|---|
| 1 | hsa-miR-140-5p | 2.71 | 11 | 61% |
| 2 | hsa-miR-146a-5p | 3.89 | 9 | 62% |
| 3 | hsa-miR-27a-3p | 3.74 | 12 | 63% |
| 4 | hsa-miR-34a-5p | 5.23 | 7 | 54% |
| 5 | hsa-miR-223-5p | 3.39 | 10 | 66% |
| 6 | hsa-miR-125b-5p | 7.29 | 8 | 69% |
| 7 | hsa-miR-675-5p | 3.32 | 5 | 54% |
| 8 | hsa-miR-101-3p | 10.25 | 12 | 74% |
| 9 | hsa-miR-148a-5p | 0.56 | 5 | 52% |
| 10 | hsa-miR-558 | 3.97 | 15 | 67% |
| 11 | hsa-miR-455-5p | 2.19 | 9 | 62% |
| 12 | hsa-miR-29a-3p | 1.05 | 5 | 53% |
| 13 | hsa-miR-145-5p | 0.73 | 6 | 54% |
| 14 | hsa-miR-149-5p | 3.25 | 5 | 64% |
| 15 | hsa-miR-26a-5p | 1.78 | 4 | 53% |
| 16 | hsa-miR-483-5p | 2.66 | 8 | 54% |
| 17 | hsa-miR-93-5p | 3.67 | 9 | 68% |
| 18 | hsa-miR-193b-5p | 2.23 | 9 | 56% |
| 19 | hsa-miR-33a-3p | 1.07 | 8 | 56% |
| 20 | hsa-miR-210-5p | 11.51 | 9 | 76% |
| 21 | hsa-miR-199a-5p | 3.35 | 12 | 63% |
| 22 | hsa-miR-9-5p | 1.97 | 9 | 62% |
| 23 | hsa-miR-355-5p | 3.95 | 10 | 62% |
| 24 | hsa-miR-22-5p | 3.28 | 7 | 54% |
| 25 | hsa-miR-24-3p | 2.19 | 13 | 66% |
| 26 | hsa-miR-21-5p | 3.29 | 10 | 69% |
| 27 | hsa-miR-98-5p | 5.25 | 5 | 54% |
| 28 | hsa-miR-127-3p | 7.39 | 15 | 74% |
| 29 | hsa-miR-488-5p | 1.91 | 7 | 52% |
| 30 | hsa-miR-181b-5p | 2.87 | 11 | 67% |
| **Average** | ———— | **3.67** | **8.8** | **61.1%** |

*1) Highly friendly user interface along with various searching, viewing, and saving options:* We use hsa-miR-140-5p as

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | microRNA | gene_symbol | gene_name | miRDB_score | TargetScan_score | miRanda_score | miRTarBase_id | pubmed_ids |
| 2 | hsa-miR-140-5p | MMD | monocyte to macrophage differentiation associated | 98.4096421 | 99 | 1.2573 | MIRT438745 | 24971538 |
| 3 | hsa-miR-140-5p | FGF9 | fibroblast growth factor 9 | 97.4421643 | 99 | 1.1093 | MIRT007378 | 23401231 |
| 4 | hsa-miR-140-5p | HDAC7 | histone deacetylase 7 | 97.29923 | 96 | 0.8225 | MIRT438365 | 24530397;26898430 |
| 5 | hsa-miR-140-5p | FOXP2 | forkhead box P2 | 91.53511 | 92 | 1.1064 | - | 24133256 |
| 6 | hsa-miR-140-5p | ADAM10 | ADAM metallopeptidase domain 10 | 87.6302 | 97 | 1.0599 | - | 27033573;26704053;24530397 |
| 7 | hsa-miR-140-5p | BMP2 | bone morphogenetic protein 2 | 78.5012 | 98 | 1.324 | - | 24928442 |
| 8 | hsa-miR-140-5p | LAMC1 | laminin subunit gamma 1 | 73.0137 | 96 | 1.0796 | MIRT438366 | 24530397;25087724 |
| 9 | hsa-miR-140-5p | TGFBR1 | transforming growth factor beta receptor I | 65.5248 | 95 | 0.1989 | MIRT007379 | 23401231;26657345 |
| 10 | hsa-miR-140-5p | VEGFA | vascular endothelial growth factor A | 64.417 | 98 | 0.9649 | MIRT003812 | 26402430;27035554 |
| 11 | hsa-miR-140-5p | RALA | v-ral simian leukemia viral oncogene homolog A (ras related) | 63.1337 | 94 | 0.9967 | MIRT437593 | 24063364 |
| 12 | hsa-miR-140-5p | PDGFRA | platelet derived growth factor receptor alpha | 61.1917 | 93 | 0.3169 | MIRT006218 | 26297547 |
| 13 | hsa-miR-140-5p | PSEN1 | presenilin 1 | 55.9715225 | - | 0.4694 | - | 24530397 |
| 14 | hsa-miR-140-5p | TP63 | tumor protein p63 | - | - | 0.4978 | - | 26695686 |
| 15 | hsa-miR-140-5p | PIN1 | peptidylprolyl cis/trans isomerase, NIMA-interacting 1 | - | 97 | 1.0593 | - | 26150353 |
| 16 | hsa-miR-140-5p | SOX9 | SRY-box 9 | - | - | - | MIRT053113 | 24063364;26175215 |
| 17 | hsa-miR-140-5p | ERBB4 | erb-b2 receptor tyrosine kinase 4 | - | 84 | - | - | 24530397 |
| 18 | hsa-miR-140-5p | STAT1 | signal transducer and activator of transcription 1 | - | - | 0.266 | - | 26780721 |
| 19 | hsa-miR-140-5p | SMAD2 | SMAD family member 2 | - | 59 | - | - | 22143896;25980495 |
| 20 | hsa-miR-140-5p | MEG3 | maternally expressed 3 (non-protein coding) | - | - | 0.9809 | - | 26898430 |
| 21 | hsa-miR-140-5p | IGFBP5 | insulin like growth factor binding protein 5 | - | 93 | 0.4617 | MIRT006556 | 25983620 |
| 22 | hsa-miR-140-5p | TAL1 | T-cell acute lymphocytic leukemia 1 | - | - | 0.5877 | - | 26882564 |
| 23 | hsa-miR-140-5p | EGFR | epidermal growth factor receptor | - | - | 0.1754 | - | 24530397;23098991 |
| 24 | hsa-miR-140-5p | DNMT1 | DNA (cytosine-5-)-methyltransferase 1 | - | - | 0.1849 | MIRT006890 | 27021683 |
| 25 | hsa-miR-140-5p | PAX6 | paired box 6 | - | 38 | 0.1592 | MIRT438364 | 24530397 |
| 26 | hsa-miR-140-5p | IGF1R | insulin like growth factor 1 receptor | - | 49 | - | MIRT054456 | 24530397 |

Fig. 3. Actual contents of the saved search results on hsa-miR-140-5p ("Query_Results_for_hsa-miR-140-5p-2016-08-17.csv").

an example to demonstrate the friendly user interface design and convenient options in the OmniSearch software tool.

- Inside the filters panel in the user interface, appropriate radio buttons can be freely selected and applied according to three filtering criteria: "Data Source Filter," "Validation Filter," and "Publications Filter." This feature allows users to customize their search results based on unique user needs, and to view these results from various, customizable aspects as well.
- Search results can be sorted on all columns in the result table, e.g., by different ranking scores from various prediction databases or by PMIDs, thereby further facilitating users to discover results of interest.
- Additional analysis can be conducted on retrieved results, including DAVID and PANTHER analysis (each having numerous analyzing options).
- Upon completion of each search query answering, users have the choice to save the search results for all or customized rows, and, in two different formats: tab-delimited text format or comma-separated values (CSV) format. Note that self-explanatory file names are automatically generated for users' convenience, "Target_List_for_hsa-miR-140-5p-2016-08-17.txt" for example. Fig. 3 shows the actual contents of the saved file of "Query_Results_for_hsa-miR-140-5p-2016-08-17.csv."

*2) Effective and efficient search results:* As demonstrated in our earlier studies [15] [22], the OmniSearch software tool is able to retrieve results in an effective and efficient manner. That is, OmniSearch answers search queries with correct (precise) results, and, in a more efficient manner compared to conventional search methods.

Our experimental results reported in this paper show that:

- OmniSearch returned meaningful results that were **not** obtained in manual search. For example, compared with the results by manual search, eight more predicted target genes were obtained by OmniSearch for hsa-miR-146a-5p: RAC1, MSC, TLR4, SIKE1, CXCL12, CXCL8, RARB, CDKN3, and EGFR. Statistics on different numbers of putative target genes returned by manual search and OmniSearch, respectively, is exhibited in Table II. Note that *all putative genes returned from the OmniSearch interface were supported by one or more PubMed publications*. These results were also verified by two domain experts, B. Liu (Department of Orthopedics, Shenzhen Luohu People's Hospital) and Y. Liu (Department of Orthopedics, Shenzhen People's Hospital). For example, the eight additional target genes listed above were supported by the following publications: PMID 25214035, PMID 23963696, PMIDs 26997759 and 27029214, PMID 23963696, PMID 23963696, PMID 26859141, PMID 27011326, PMID 26859141, and PMIDs 26697527 and 26857280, respectively.
- OmniSearch was much more efficient than conventional search. We asked the aforementioned two domain experts to perform queries on 30 miRs using their conventional methods; next, they performed the same set of queries through the OmniSearch interface; and finally, the saved time for both domain experts were averaged and exhibited in Table III.

## V. Conclusions

Osteoarthritis (OA) is the most common joint disease worldwide, resulting in severe joint pain and significantly decreased life quality in the elderly population. Due to the lack of effective medicine that can reverse the degeneration of articular cartilage, ultimately all OA patients need to receive an artificial joint replacement surgery. Although the pathogenesis and progression of OA has attracted large amounts of research activities, most regulation mechanisms at genetic and genomics level have not yet been completely understood. In this paper, we introduce the utilization of OmniSearch, a semantic integration and search software tool, to facilitate the exploration

of important roles performed by different miRs on OA disease process. We have obtained experimental results with good performance, in particular, regarding the software's effectiveness (accuracy) and efficiency. Our methodologies are demonstrated to be promising in unraveling miR regulation mechanisms in OA. As a result, the utilization of OmniSearch can help better understand the OA pathogenesis and development at genetic level, which, in turn, will make the early diagnosis of OA possible. It can also facilitate clinical investigators to explore more effective treatment for this serious disease in the future.

We plan to conduct wet-lab experiments in our future work, including real-time quantitative PCR (qPCR) and dual luciferase reporter assay, among others, to validate biological functions of computationally putative miR::target gene pairs output from the OmniSearch system.

## ACKNOWLEDGMENT

Research reported in this paper was partially supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH), under the Award Number U01CA180982. The views contained in this paper are solely the responsibility of the authors and do not represent the official views, either expressed or implied, of the NIH or the U.S. Government.

## REFERENCES

[1] D. Iliopoulos, K. Malizos, and P. Oikonomou, "Integrative microRNA and proteomic approaches identify novel osteoarthritis genes and their collaborative metabolic and inflammatory networks," *PLOS ONE*, vol. 11, pp. 3740–3750, 2008.

[2] S. Jones, G. Watkins, and N. L. Good, "The identification of differentially expressed microRNA in osteoarthritic tissue that modulate the production of TNF-alpha and MMP13," *Osteoarthritis Cartilage*, vol. 4, pp. 464–472, 2009.

[3] E. Wienholds, W. Kloosterman, and E. Miska, "MicroRNA expression in zebrafish embryonic development," *Science*, vol. 309, no. 5732, pp. 310–311, 2005.

[4] L. Tuddenham, G. Wheeler, and S. Ntounia-Fousara, "The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells," *FEBS Lett*, vol. 580, no. 17, pp. 4214–4217, 2006.

[5] S. Miyaki, T. Nakasa, and S. Otsuki, "MicroRNA-140 is expressed in differentiated human articular chondrocytes and modulates interleukin-1 responses," *Arthritis Rheum*, vol. 60, no. 9, pp. 2723–2730, 2009.

[6] T. Matsukawa, T. Sakai, and T. Yonezawa, "MicroRNA-125b regulates the expression of aggrecanase-1 (ADAMTS-4) in human osteoarthritic chondrocytes," *Arthritis Res Ther*, vol. 15, no. 1, 2013.

[7] L. Vonk, A. Kragten, and W. Dhert, "Overexpression of hsa-miR-148a promotes cartilage production and inhibits cartilage degradation by osteoarthritic chondrocytes," *Osteoarthritis Cartilage*, vol. 22, no. 1, pp. 145–153, 2014.

[8] D. Zhang, X. Cao, and J. Li, "MiR-210 inhibits NF-B signaling pathway by targeting DR6 in osteoarthritis," *Sci Rep*, vol. 5, no. 12775, 2015.

[9] M. Ashburner, C. Ball, J. Blake, D. Botstein, H. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, M. Harris, D. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. Matese, J. Richardson, M. Ringwald, G. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000.

[10] J. Huang, K. Eilbeck, B. Smith, J. Blake, D. Dou, W. Huang, D. Natale, A. Ruttenberg, J. Huan, M. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. Strachan, N. de Silva, M. Kasukurthi, V. Jha, Y. He, S. Zhang, X. Wang, Z. Liu, G. Borchert, , and M. Tan, "The Development of Non-Coding RNA Ontology," *Int. J. Data Mining and Bioinformatics*, vol. 15, no. 3, pp. 214–232, June 2016.

[11] J. Huang, K. Eilbeck, B. Smith, J. Blake, D. Dou, W. Huang, D. Natale, A. Ruttenberg, J. Huan, M. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. Strachan, Y. He, S. Zhang, X. Wang, Z. Liu, G. Borchert, , and M. Tan, "The Non-Coding RNA Ontology (NCRO): A comprehensive resource for the unification of non-coding RNA biology," *J Biomed Semantics*, vol. 7, no. 24, May 2016.

[12] OBO Library. [Online]. Available: http://obofoundry.org

[13] J. Huang, J. Dang, X. Lu, D. Dou, J. Blake, W. Gerthoffer, and M. Tan, "An Ontology-Based MicroRNA Knowledge Sharing and Acquisition Framework," in *Proc. BHI Workshop at 2012 IEEE Intl' Conference on Bioinformatics and Biomedicine (BIBM-2012)*, pp. 16–23.

[14] J. Huang, J. Dang, G. Borchert, K. Eilbeck, H. Zhang, M. Xiong, W. Jiang, H. Wu, J. Blake, D. Natale, and M. Tan, "OMIT: Dynamic, Semi-Automated Ontology Development for the microRNA Domain," *PLOS ONE*, vol. 9, no. 7, July 2014.

[15] J. Huang, F. Gutierrez, H. Strachan, D. Dou, W. Huang, B. Smith, J. Blake, K. Eilbeck, D. Natale, Y. Lin, B. Wu, N. de Silva, X. Wang, Z. Liu, G. Borchert, M. Tan, and A. Ruttenberg, "OmniSearch: A semantic search system based on the Ontology for MIcroRNA Target (OMIT) for microRNA-target gene interaction data," *J Biomed Semantics*, vol. 7, no. 25, May 2016.

[16] D. Dou, H. Wang, and H. Liu, "Semantic Data Mining: A Survey of Ontology-based Approaches," in *Proc. IEEE International Conference on Semantic Computing*, 2015, pp. 244–251.

[17] W. Premerlani and M. Blaha, "An Approach for Reverse Engineering of Relational Databases," *Commun. ACM Journal*, vol. 37, no. 5, pp. 42–49, 1994.

[18] L. Stojanovic, N. Stojanovic, and R. Volz, "Migrating data-intensive web sites into the Semantic Web," in *Proc. ACM symposium on Applied computing*, 2002, pp. 1100–1107.

[19] P. Verheyden, J. Bo, and R. Meersman, "Semantically Unlocking Database Content Through Ontology-Based Mediation," in *Proc. SWDB 2004*, 2004, pp. 109–126.

[20] L. Lubyte and S. Tessaris, "Extracting Ontologies from Relational Databases," in *Proc. 2007 Description Logics*, 2007, pp. 122–126.

[21] R. Chauhan, R. Gouda, R. Sharma, and A. Chauhan, "Domain ontology based semantic search for efficient information retrieval through automatic query expansion," in *Proc. 2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*, March 2013, pp. 397–402.

[22] J. Huang, F. Gutierrez, D. Dou, J. Blake, K. Eilbeck, D. Natale, B. Smith, Y. Lin, X. Wang, Z. Liu, M. Tan, and A. Ruttenberg, "A semantic approach for knowledge capture of microRNA-target gene interactions," in *Proc. BHI Workshop at 2015 IEEE Intl' Conference on Bioinformatics and Biomedicine (BIBM-2015)*, pp. 975–982.

[23] Apache. [Online]. Available: https://jena.apache.org/

[24] SPARQL Query Language for RDF. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/

[25] RDF: Resource Description Framework. [Online]. Available: http://www.w3.org/RDF/

[26] miRDB. [Online]. Available: http://mirdb.org/miRDB/

[27] TargetScan. [Online]. Available: http://www.targetscan.org

[28] miRanda. [Online]. Available: http://www.microrna.org

[29] miRTarBase. [Online]. Available: http://mirtarbase.mbc.nctu.edu.tw/

[30] miRBase. [Online]. Available: http://www.mirbase.org/

[31] PubMed. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed

[32] NCBI Gene. [Online]. Available: http://www.ncbi.nlm.nih.gov/gene

[33] MeSH. [Online]. Available: http://www.ncbi.nlm.nih.gov/mesh

[34] Gene Ontology Annotation (GOA) Database. [Online]. Available: http://geneontology.org/page/download-annotations

[35] RNAcentral. [Online]. Available: http://rnacentral.org