



NIH NCI ITCR Annual PI Meeting

University of California, Santa Cruz, May 31 – June 2,
2017

Presented by:
Jingshan Huang
Biomedical Informatics Group
University of South Alabama



Presentation Outline

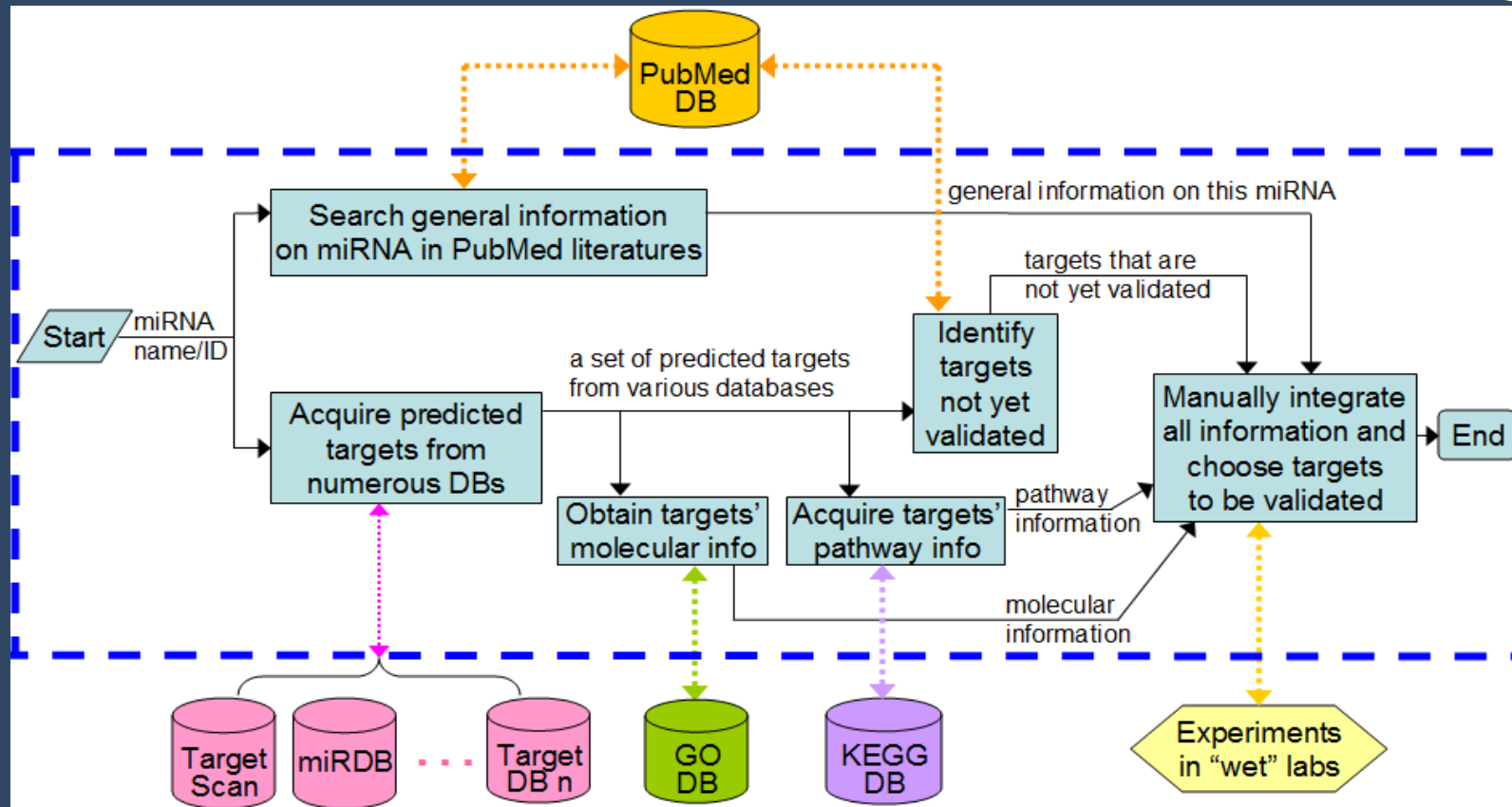
- Project overview
- Progress report
- Software evaluation



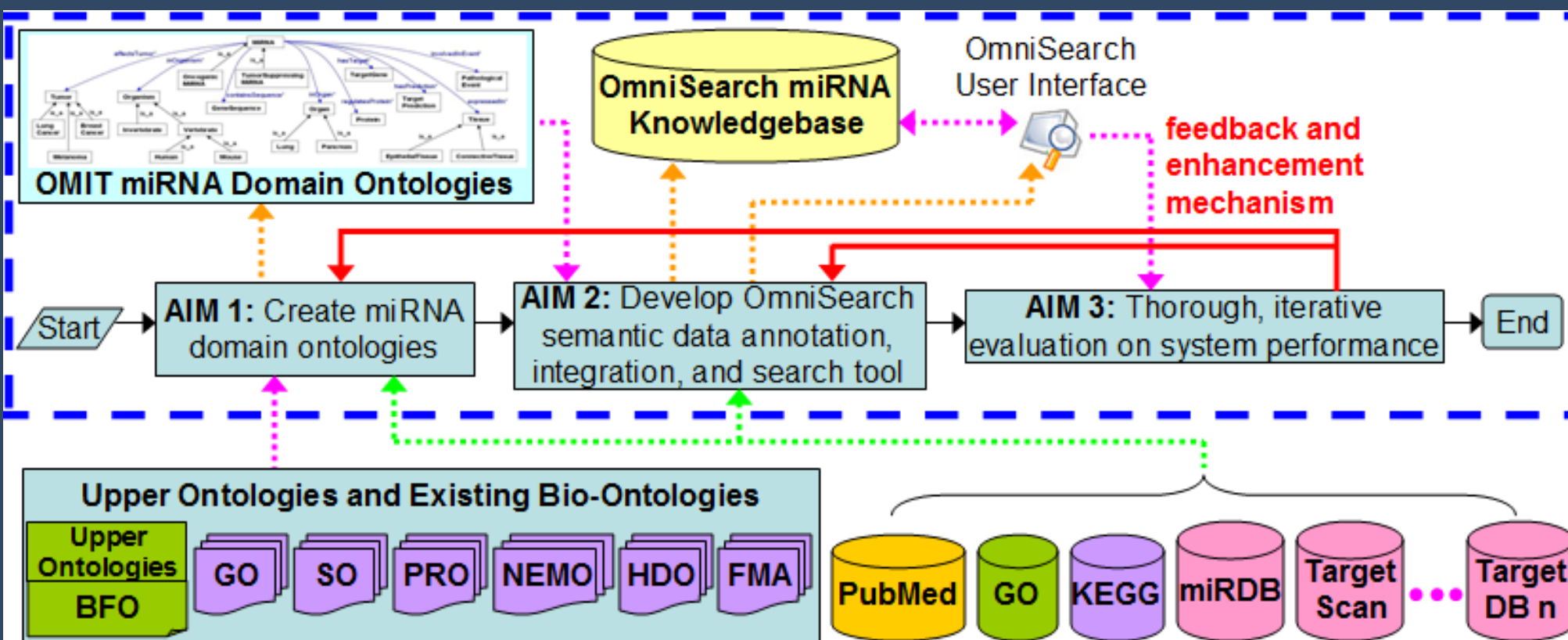
Project Overview

- OmniSearch is a semantic integration and search software tool specifically designed for microRNA (miR) cancer research
- It will assist cancer researchers in unraveling critical roles of miR regulations in human cancers
- Targeted users: cancer biologists, bioinformaticians, and clinical investigators
- OmniSearch can be used to obtain unified knowledge related to miR regulation mechanisms and thus derive unique insights for the regulation and control of cancer disease

A typical workflow demonstrating current search approaches adopted by many cancer researchers



Notion: manual integration of information from heterogeneous sources has become labor-intensive and error-prone





Project Timeline

	Project Activities	Milestones by the End of Each Year
Y1	<ul style="list-style-type: none">• Start ontology development• Start data annotation & integration	<ul style="list-style-type: none">• Initial version of OMIT ontologies• “Key Phrase Extraction” & “Ontology Mapping” modules
Y2	<ul style="list-style-type: none">• Continue ontology development• Continue data annotation & integration	<ul style="list-style-type: none">• Stable version of OMIT ontologies• New terminology contributed to bio-ontology community• Other software modules; initial OmniSearch package
Y3	<ul style="list-style-type: none">• Design a friendly semantic search GUI• Develop use cases and evaluating queries• Conduct iterative system evaluation• Integrate feedback from the community	<ul style="list-style-type: none">• Final version of OMIT ontologies and OmniSearch package• A comprehensive miRNA KB unified from numerous sources• A set of semantic search evaluating queries• A collection of formative evaluation reports• A final, summative evaluation report



Project Timeline

	Project Activities	Milestones by the End of Each Year
Y1	<ul style="list-style-type: none">• Start ontology development• Start data annotation & integration	<ul style="list-style-type: none">• Initial version of OMIT ontologies• “Key Phrase Extraction” & “Ontology Mapping” modules
Y2	<ul style="list-style-type: none">• Continue ontology development• Continue data annotation & integration	<ul style="list-style-type: none">• Stable version of OMIT ontologies• New terminology contributed to bio-ontology community• Other software modules; initial OmniSearch package
Y3	<ul style="list-style-type: none">• Design a friendly semantic search GUI• Develop evaluating queries• Conduct iterative system evaluation• Integrate feedback from the community	<ul style="list-style-type: none">• Final version of OMIT ontologies and OmniSearch package• A comprehensive miRNA KB unified from numerous sources• A set of semantic search evaluating queries• A collection of formative evaluation reports• A final, summative evaluation report

- **OMIT: OBO Foundry/NCBO BioPortal/GitHub**
- **OmniSearch: Browser-based/GitHub**
- **KB: 400 million RDF triples**
- **115 evaluating queries**



Project Timeline

	Project Activities	Milestones by the End of Each Year
Y1	<ul style="list-style-type: none">• Start ontology development• Start data annotation & integration	<ul style="list-style-type: none">• Initial version of OMIT ontologies• “Key Phrase Extraction” & “Ontology Mapping” modules
Y2	<ul style="list-style-type: none">• Continue ontology development• Continue data annotation & integration	<ul style="list-style-type: none">• Stable version of OMIT ontologies• New terminology contributed to bio-ontology community• Other software modules; initial OmniSearch package
Y3	<ul style="list-style-type: none">• Design a friendly semantic search GUI• Develop evaluating queries• Conduct iterative system evaluation• Integrate feedback from the community	<ul style="list-style-type: none">• Final version of OMIT ontologies and OmniSearch package• A comprehensive miRNA KB unified from numerous sources• A set of semantic search evaluating queries• A collection of formative evaluation reports• A final, summative evaluation report
Y4	<ul style="list-style-type: none">• Conduct iterative system evaluation• Integrate feedback from the community	<ul style="list-style-type: none">• A collection of formative evaluation reports• A final, summative evaluation report



Project Dissemination

- **Collaborating with a wide range of bio-ontology community:**
Gene Ontology (GO), Sequence Ontology (SO), PRotein Ontology (PRO),
Chemical Entities of Biological Interest Ontology (CHEBI), Ontology for
Biomedical Investigations (OBI), and Uber Anatomy Ontology (UBERON)



Project Dissemination

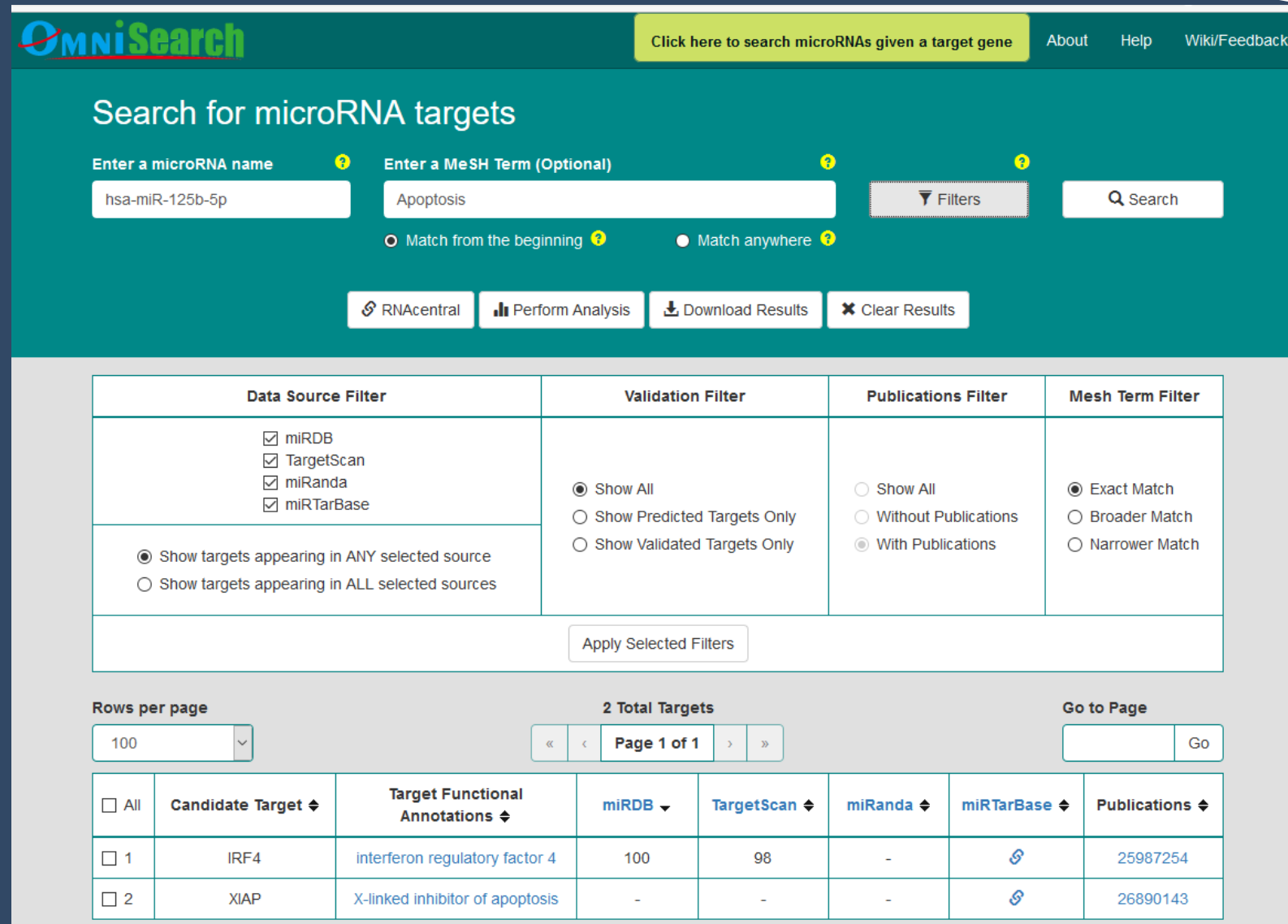
- **Collaborating with a wide range of bio-ontology community:**
Gene Ontology (GO), Sequence Ontology (SO), PRotein Ontology (PRO), Chemical Entities of Biological Interest Ontology (CHEBI), Ontology for Biomedical Investigations (OBI), and Uber Anatomy Ontology (UBERON)
- **OBO Foundry:** <http://www.obofoundry.org/ontology/omit.html>
- **NCBO BioPortal:** <https://bioportal.bioontology.org/ontologies/OMIT>
- **Project wiki site:** <http://omnisearch.soc.southalabama.edu/>
- **GitHub:** <https://github.com/OmniSearch>



Project Dissemination

- **Collaborating with a wide range of bio-ontology community:**
Gene Ontology (GO), Sequence Ontology (SO), PRotein Ontology (PRO), Chemical Entities of Biological Interest Ontology (CHEBI), Ontology for Biomedical Investigations (OBI), and Uber Anatomy Ontology (UBERON)
- **OBO Foundry:** <http://www.obofoundry.org/ontology/omit.html>
- **NCBO BioPortal:** <https://bioportal.bioontology.org/ontologies/OMIT>
- **Project wiki site:** <http://omnisearch.soc.southalabama.edu/>
- **GitHub:** <https://github.com/OmniSearch>
- **GUI Tutorial on YouTube:** <https://www.youtube.com/watch?v=kCFm4YkNvEg>

Software Interface (1)



The screenshot shows the OmniSearch web interface. At the top, there is a navigation bar with the OmniSearch logo, a link to search microRNAs given a target gene, and links for About, Help, and Wiki/Feedback. The main heading is "Search for microRNA targets". Below this, there are input fields for "Enter a microRNA name" (containing "hsa-miR-125b-5p") and "Enter a MESH Term (Optional)" (containing "Apoptosis"). There are also radio buttons for "Match from the beginning" and "Match anywhere". A "Filters" button and a "Search" button are present. Below the search area, there are buttons for "RNAcentral", "Perform Analysis", "Download Results", and "Clear Results".

The filter section contains four columns of options:

- Data Source Filter:**
 - miRDB
 - TargetScan
 - miRanda
 - miRTarBase
 - Show targets appearing in ANY selected source
 - Show targets appearing in ALL selected sources
- Validation Filter:**
 - Show All
 - Show Predicted Targets Only
 - Show Validated Targets Only
- Publications Filter:**
 - Show All
 - Without Publications
 - With Publications
- Mesh Term Filter:**
 - Exact Match
 - Broader Match
 - Narrower Match

An "Apply Selected Filters" button is located below the filter section.

Below the filters, there is a "Rows per page" dropdown set to 100, a "2 Total Targets" indicator, and a "Go to Page" field with a "Go" button. The results are displayed in a table with the following columns: Candidate Target, Target Functional Annotations, miRDB, TargetScan, miRanda, miRTarBase, and Publications.

<input type="checkbox"/> All	Candidate Target	Target Functional Annotations	miRDB	TargetScan	miRanda	miRTarBase	Publications
<input type="checkbox"/> 1	IRF4	interferon regulatory factor 4	100	98	-	🔗	25987254
<input type="checkbox"/> 2	XIAP	X-linked inhibitor of apoptosis	-	-	-	🔗	26890143

Software Interface (2)

Search for microRNAs given a target gene

Enter a Target gene [?]

Enter a microRNA (Optional) [?]

Enter a MeSH Term (Optional) [?]

[?]

[?]

Match from the beginning [?]

Match anywhere [?]

Rows per page

130 Total microRNAs

« < Page 1 of 2 > »

Go to Page

<input type="checkbox"/> All	Targeting microRNA [▲]	RNAcentral	miRDB [◆]	TargetScan [◆]	miRanda [◆]	miRTarBase [◆]	Publications [◆]
<input type="checkbox"/> 1	hsa-miR-103a-2-5p	🔗	80	-	-	-	-
<input type="checkbox"/> 2	hsa-miR-103a-3p	🔗	54	-	-	-	-
<input type="checkbox"/> 3	hsa-miR-1178-5p	🔗	60	-	-	-	-
<input type="checkbox"/> 4	hsa-miR-1199-5p	🔗	-	-	-	🔗	-
<input type="checkbox"/> 5	hsa-miR-1225-3p	🔗	-	-	-	🔗	-
<input type="checkbox"/> 6	hsa-miR-1225-5p	🔗	62	-	-	-	-
<input type="checkbox"/> 7	hsa-miR-1233-3p	🔗	-	-	-	🔗	-
<input type="checkbox"/> 8	hsa-miR-1251-3p	🔗	59	-	-	-	-
<input type="checkbox"/> 9	hsa-miR-125a-5p	🔗	100	98	0 1239	🔗	-



Wiki-Style Project Site

OmniSearch

Main page Discussion Read View source View history Search

Main Page

[Main Page](#) | [Project Members](#) | [Collaborating Labs](#) | [Ontology](#) | [Software](#) | [Publications](#) | [Contact Us](#)

Welcome to the OmniSearch Project

Funded by National Institutes of Health (NIH)/NCI 1U01CA180982-01A1 under ITCR Initiative

microRNAs (short for miRNAs or miRs) are small, non-coding RNA molecules that regulate the expression of their respective target proteins. As such, miRNAs are associated with the development, diagnosis, treatment, and prognosis of various human diseases. The complete elucidation of miRNA interactions requires more than a list of biologically validated miRNA targets, as these could represent a subset of the possibly hundreds of targets. miRNA target prediction databases can fill potential gaps. However, the ~30 such databases use different prediction algorithms and use heterogeneous semantics, preventing easy comparison. In addition, gene expression, protein function, and impacted processes are critical pieces of information for biologists to determine the impact of miRNAs. OmniSearch software was developed to handle the significant challenge of miRNA-related data integration and query. It will significantly assist biologists, bioinformaticians, and clinical investigators to unravel critical roles performed by different microRNAs in human disease.

Start using [OmniSearch UI here](#), and the source code can be accessed [here](#).

Before posting a new comment, browse [Recent Discussions](#) or search existing discussions below.

Email Author: huang@southalabama.edu

Main Page
Project Members
Collaborating Labs
Ontology
Software
Publications
Latest News
Contact Us
Help
Copyrights
Tools
[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)
[Page information](#)



Evaluation via Google Docs

Signing-up sheet for organizers:

	A	B	C	D	E	F	G
	Evaluator Number	Evaluator Name (First + Last Initial)	Assigned Queries	Training or not (Y/N)?	Completed or not (Y/N)?	Institution	Group Organizer
1	1	Bernie L.	88 27 17	Y		University of Utah School of Medicine	Karen E.
2	2	Nick C.	34 67 23	Y	Y	University of Utah School of Medicine	Karen E.
3	3	Lance P.	89 55 114	Y		University of Utah School of Medicine	Karen E.
4	4	Andrew M.	30 100 87	N	Y	University of Utah School of Medicine	Karen E.
5	5	Uerlie K.	92 17 34	Y		University of Utah School of Medicine	Karen E.
6	6	Lesli S.	93 101 54	Y	Y	University of Utah School of Medicine	Karen E.
7	7	Jason S.	2 81 6	Y	Y	University of Utah School of Medicine	Karen E.
8	8	Aly K.	92 54 104	N	Y	University of Utah School of Medicine	Karen E.
9	9	Kort Z.	49 63 62	Y		University of Utah School of Medicine	Karen E.
10	10	Nicole R.	21 112 109	N	Y	University of Utah School of Medicine	Karen E.
11	11	Sandra C.	66 9 94	Y		University of Utah School of Medicine	Karen E.
12	12	Deb H.	76 110 105	N		University of Utah School of Medicine	Karen E.
13	13	Lisa B.	40 96 42	Y		University of Utah School of Medicine	Karen E.
14	14	Marc P.	21 22 105	N	Y	University of Utah School of Medicine	Karen E.
15	15	Oscar R.	98 76 40	Y	Y	University of Utah School of Medicine	Karen E.
16	16	Simon T.	84 47 85	N		University of Utah School of Medicine	Karen E.
17	17	Brandi G.	97 101 85	Y		University of Utah School of Medicine	Karen E.
18	18	Tom H.	45 40 8	N	Y	University of Utah School of Medicine	Karen E.
19	19	Mark Z.	12 36 87	Y		University of Utah School of Medicine	Karen E.
20	20	Karen C.	79 104 91	N	Y	University of Utah School of Medicine	Karen E.
21	21	Diane W.	66 86 91	Y		University of Utah School of Medicine	Karen E.
22	22	Na H.	24 88 21	Y	Y	Mayo Clinic College of Medicine, Division of Biomedical Statistics & Informatics	Na H.
23	23	Kui W.	83 108 11	Y	Y	Mayo Clinic College of Medicine, Division of Biomedical Statistics & Informatics	Na H.
24	24	Deepak S.	92 105 77	Y	Y	Mayo Clinic College of Medicine, Division of Biomedical Statistics & Informatics	Na H.
25	25	Vikas S.	52 30 89	N	Y	Mayo Clinic College of Medicine, Division of Biomedical Statistics & Informatics	Na H.
26	26	Jie W.	109 31 102	Y	Y	Mayo Clinic College of Medicine, Division of Biomedical Statistics & Informatics	Na H.
27	27	Yanan Y.	98 64 91	N	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
28	28	Lixia G.	85 11 83	Y	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
29	29	Chaofan L.	84 81 102	N	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
30	30	Ting Z.	114 107 83	Y	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
31	31	Bibo Z.	2 41 58	N	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
32	32	Yuanyuan W.	93 1 35	Y		Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
33	33	Zheng W.	37 49 36	Y	Y	Mayo Clinic Cancer Center, Division of Pulmonary and Critical Care Medicine	Yanan Y.
34	34	Shari T.	108 56 27	Y		Georgetown University Medical Center	Darren N.
35	35	Karen R.	101 69 111	N		Georgetown University Medical Center	Darren N.
36	36	Cecilia A.	93 41 86	Y		Georgetown University Medical Center	Darren N.
37	37	Darren N.	12 89 9	N		Georgetown University Medical Center	Darren N.
38	38		42 39 1	Y		Georgetown University Medical Center	Darren N.
39	39		1 98 10	N		Georgetown University Medical Center	Darren N.
40	40		92 4 50	Y		Georgetown University Medical Center	Darren N.
41	41	Xinsong D.	24 85 113	N	Y	University of Florida College of Medicine	Jiang B.
42	42	Yi G.	5 49 62	Y		University of Florida College of Medicine	Jiang B.
43	43	Qian L.	47 4 9	N	Y	University of Florida College of Medicine	Jiang B.
44	44	Hansi Z.	100 32 82	Y	Y	University of Florida College of Medicine	Jiang B.
45	45	Francois M.	77 7 98	N	Y	University of Florida College of Medicine	Jiang B.
46	46	Yunpeng Z.	64 51 39	Y	Y	University of Florida College of Medicine	Jiang B.
47	47	Jiang B.	110 66 17	N	Y	University of Florida College of Medicine	Jiang B.
48	48	He, Y.	66 70 23	Y		University of Michigan Medical School	Yongqun H.
49	49	Ong, E.	59 82 110	N		University of Michigan Medical School	Yongqun H.
50	50	Wang, H.	76 2 51	Y		University of Michigan Medical School	Yongqun H.
51	51	Yu, H.	36 61 55	N		University of Michigan Medical School	Yongqun H.
52	52	Wong, M.	110 104 18	Y		University of Michigan Medical School	Yongqun H.

- A total of 121 evaluators
- From 19 institutions nationwide



Evaluation via Google Docs

Instructions for evaluators:

https://docs.google.com/document/d/12TFKAV7Mn0BEYfPnI_DL-4S8Hmjedes63H8GKqliuS0/

Software tutorial:

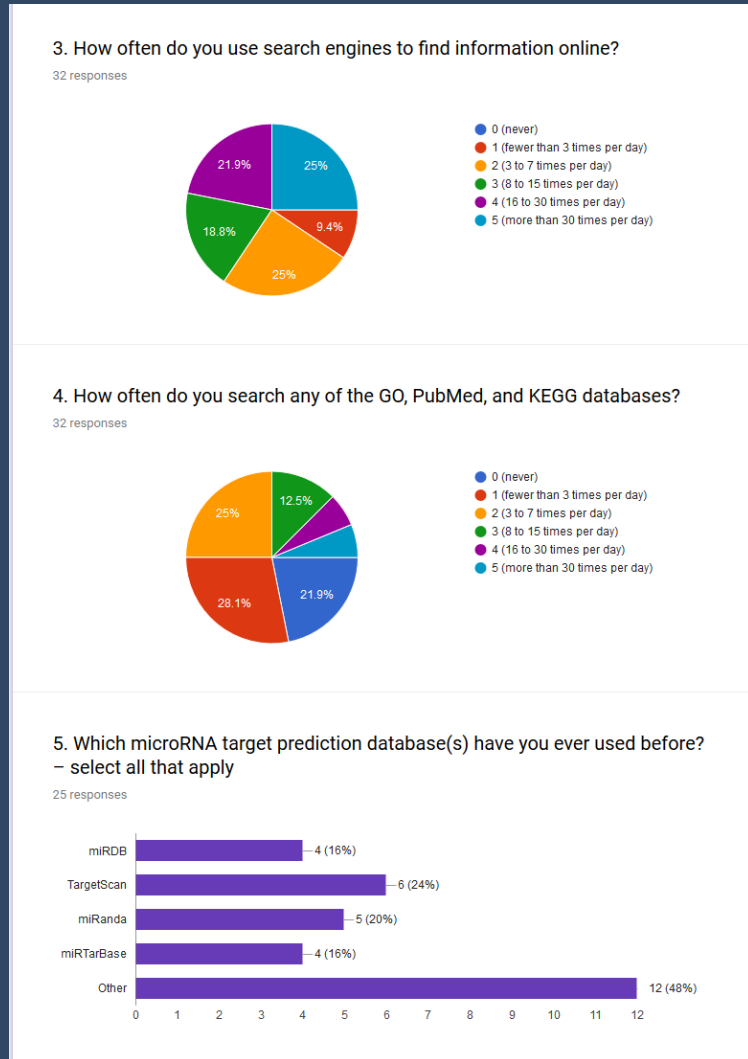
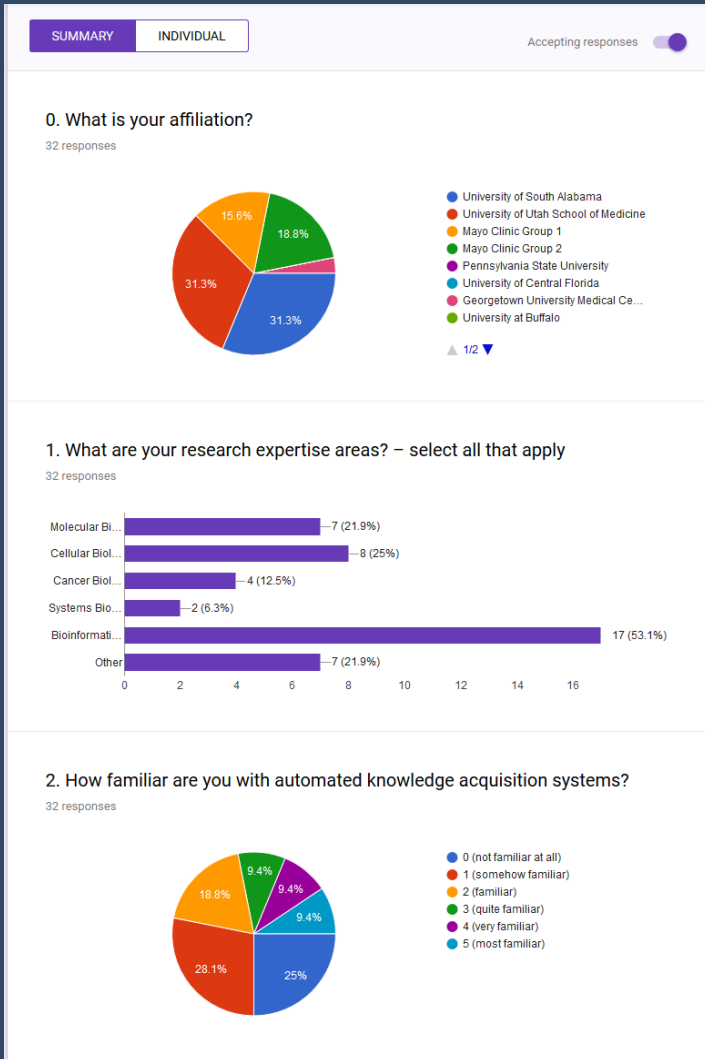
<https://www.youtube.com/watch?v=kCFm4YkNvEg>

Evaluating queries:

https://docs.google.com/document/d/1binpkOzsHYzk2gk2UDwOI2o-jaOEdD4qgr4cZil_sYk/pub

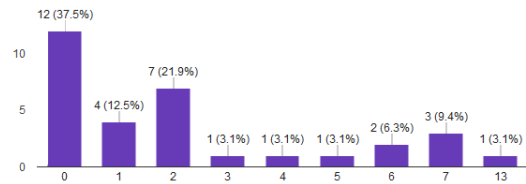
Questionnaire for evaluators:

<https://docs.google.com/forms/d/e/1FAIpQLScQ377s0DL4LYSAujtq0OylAvEYTE5SGrsNAZ5rP72TtnTV2w/viewform>



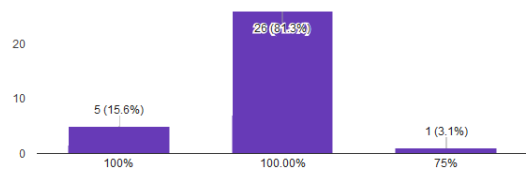
Regarding Query 2 Number of all rows (y)

32 responses



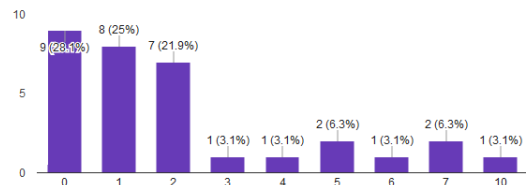
Regarding Query 2 Correctness of results (rounded to the hundredth digit) $x/y*100\%$. Note: If you get 0/0, the result should be 100.00%.

32 responses



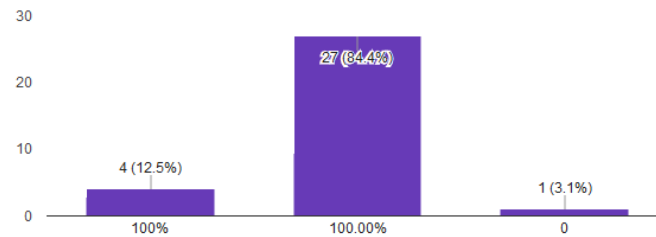
Regarding Query 3 Number of correct rows (x)

32 responses



Regarding Query 3 Correctness of results (rounded to the hundredth digit) $x/y*100\%$. Note: If you get 0/0, the result should be 100.00%.

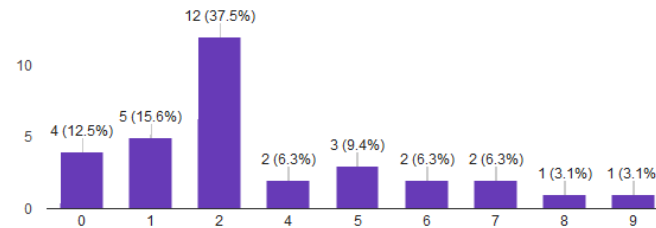
32 responses



7. List missing knowledge, i.e., any rows that you think should be, but were not, returned in the result table. Note that the value for "x" in Question #7 should be the same as the value for "x" in Question #6.

Regarding query 1 Number of correct rows (x)

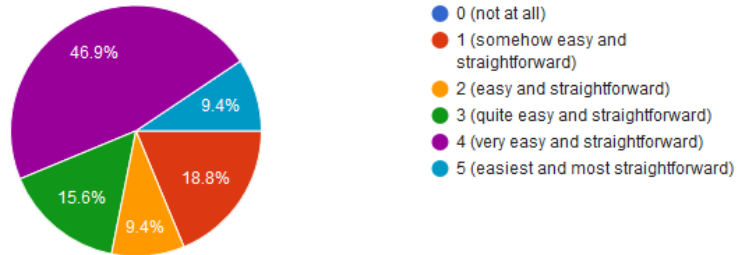
32 responses



Questions on GUI Friendliness and Usability

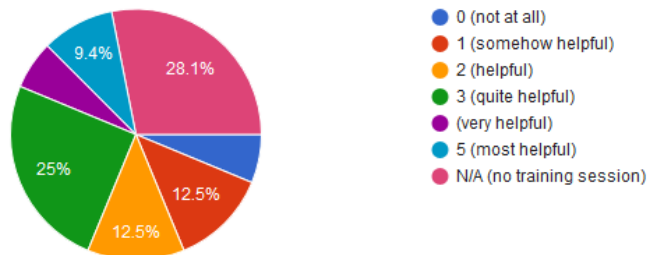
9. Is the GUI easy and straightforward to use?

32 responses



10. How helpful is the training session?

32 responses





Evaluation via Google Docs

Promising evaluation outcomes



Evaluation via Google Docs

Promising evaluation outcomes

(1) Software effectiveness and accuracy

on average, the correctness and completeness of search results were 89% and 77%, respectively



Evaluation via Google Docs

Promising evaluation outcomes

(1) Software effectiveness and accuracy

on average, the correctness and completeness of search results were 89% and 77%, respectively

(2) Software efficiency

the saved human labor was averaged at 63%

Promising evaluation outcomes

(1) Software effectiveness and accuracy

on average, the correctness and completeness of search results were 93% and 87%, respectively

(2) Software efficiency

the saved human labor was averaged at 63%

(3) Software friendliness and usability

47% of users rated the software tool as “easiest and most straightforward” (i.e., a scale of 5 out of 5); altogether 71% of users rated a scale of 4 or above out of 5



Recent Publications

So far, a total of 15 joint publications in the OmniSearch project



Recent Publications

So far, a total of 15 joint publications in the OmniSearch project

In Project Year 3, we had seven joint publications



Recent Publications

So far, a total of 15 joint publications in the OmniSearch project

In Project Year 3, we had seven joint publications

7) L. Zhang, R. Li, J. He, Q. Yang, Y. Wu, J. Huang, and B. Wu, "Co-expression analysis among microRNAs, long non-coding RNAs, and messenger RNAs to understand the pathogenesis and progression of diabetic kidney disease at the genetic level," (in press) *J. Methods. Special Issue: Integrative Analysis of Omics Data*, (2017).

6) J. Huang, G.M. Borchert, D. Dou, J. Huan, W. Lan, M. Tan, and B. Wu, editors, *Bioinformatics in microRNA research*, (in press) Springer series of Methods in Molecular Biology (2017).

5) J. Huang, D. Dou, J. She, A.H. Limper, Y. Yang, and P. Yang, "A comprehensive (biological and computational) investigation on the role of microRNA::mRNA regulations performed in chronic obstructive pulmonary disease and lung cancer," *Proc. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-16)*, pp. 1067-1072, Dec. (2016).

4) L. Zhang, R. Li, Q. Yang, Y. Wu, J. Huang, and B. Wu, "Innovative microRNA-lncRNA-mRNA co-expression analysis to understand the pathogenesis and progression of diabetic kidney disease," *Proc. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-16)*, pp. 688-693, Dec. (2016).

3) J. Huang, B. Liu, Y. Liu, and J. Chen, "The utilization of the OmniSearch semantic search tool to explore various microRNA regulation mechanisms in osteoarthritis," *Proc. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM-16)*, pp. 1073-1078, Dec. (2016).

2) J. Huang, F. Gutierrez, H. Strachan, D. Dou, W. Huang, B. Smith, J.A. Blake, K. Eilbeck, D.A. Natale, Y. Lin, B. Wu, N. de Silva, X. Wang, Z. Liu, G.M. 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.*, 2016 May 10;7:25. doi: 10.1186/s13326-016-0064-2. eCollection 2016. PubMed PMID: 27175225; PubMed Central PMCID: PMC4863347 (2016).

1) J. Huang, K. Eilbeck, B. Smith, J.A. Blake, D. Dou, W. Huang, D.A. Natale, A. Ruttenberg, J. Huan, M.T. Zimmermann, G. Jiang, Y. Lin, B. Wu, H. Strachan, Y. He, S. Zhang, X. Wang, Z. Liu, G.M. Borchert, and M. Tan, "The Non-Coding RNA Ontology (NCRO): A comprehensive resource for the unification of non-coding RNA biology," *J Biomed Semantics.*, 2016 May 4;7:24. doi: 10.1186/s13326-016-0066-0. eCollection 2016. PubMed PMID: 27152146; PubMed Central PMCID: PMC4857245 (2016).

- **NIH NCI ITCR Initiative**

- **University of South Alabama Team**
- **University of Oregon Team**
- **University of Buffalo – SUNY Team**

- **Gene Ontology**
- **Sequence Ontology**
- **PRotein Ontology**

Questions?



THANKS!