Pharm Res (2011) 28:3101–3104
DOI 10.1007/s11095-011-0573-8

COMMENTARY

OMIT: A Domain-Specific Knowledge Base for MicroRNA Target Prediction

Jingshan Huang • Christopher Townsend • Dejing Dou • Haishan Liu • Ming Tan

Received: 1 April 2011 / Accepted: 15 August 2011 / Published online: 31 August 2011

ABSTRACT Identification and characterization of the important roles microRNAs (miRNAs) perform in human cancer is an increasingly active research area. Unfortunately, prediction of miRNA target genes remains a challenging task to cancer researchers. Current processes are time-consuming, error-prone, and subject to biologists' limited prior knowledge. Therefore, we propose a domain-specific knowledge base built upon Ontology for MicroRNA Targets (OMIT) to facilitate knowledge acquisition in miRNA target gene prediction. We describe the ontology design, semantic annotation and data integration, and user-friendly interface and conclude that the OMIT system can assist biologists in unraveling the important roles of miRNAs in human cancer. Thus, it will help clinicians make sound decisions when treating cancer patients.

Electronic supplementary material The online version of this article (doi:10.1007/s11095-011-0573-8) contains supplementary material, which is available to authorized users.

J. Huang (✉) • C. Townsend
School of Computer and Information Sciences
University of South Alabama
307 University Blvd. N
Mobile, Alabama, USA
e-mail: huang@usouthal.edu

D. Dou • H. Liu
Department of Computer and Information Science
University of Oregon
Eugene, Oregon, USA

M. Tan
Mitchell Cancer Institute University of South Alabama
Mobile, Alabama, USA

M. Tan
Department of Cell Biology and Neuroscience
University of South Alabama
Mobile, Alabama, USA

KEY WORDS human cancer · knowledge acquisition · knowledge base · microRNA (miRNA) target · ontology

INTRODUCTION

The identification and characterization of the important roles microRNAs (miRNAs) perform in human cancer is an increasingly active research area. As a special class of small non-coding RNAs, miRNAs have been reported to perform critical roles in a variety of biological processes by regulating target genes (1,2). Moreover, miRNA expression profiling of many tumor types has identified miRNAs associated with cancer development, diagnosis, treatment, and prognosis (3,4). Unfortunately, the prediction of miRNA target genes remains a challenging task to cancer researchers. In particular, substantial time and effort have been expended in every search for available information in each small miRNA subarea. To identify miRNAs' target genes is very difficult: not only do biologists need to extract a large number of candidate target genes from existing miRNA target prediction databases, but they will also need to manually search for these genes' related information (e.g., their cellular components and biological processes) from resources other than miRNA databases for each of the hundreds of candidate target genes. The whole process is time-consuming, error-prone, and subject to biologists' limited prior knowledge. In addition, the situation is further aggravated by the great complexity and imprecise terminologies that characterize the biological and biomedical research fields. A great deal of variety has been identified in the adoption of different biological terms, along with divergent relationships among all these terms. Such variety has inhibited effective information acquisition by humans.

OMIT FRAMEWORK

Ontologies are formal, declarative knowledge representation models, performing a key role in defining formal semantics in traditional knowledge engineering. Therefore, we explore a domain-specific knowledge base built upon the Ontology for MicroRNA Targets (OMIT) to handle challenges in miRNA target acquisition. The OMIT ontology is *the very first ontology* in the miRNA area, and the OMIT framework facilitates knowledge discovery and sharing from existing sources. As a result, the long-term objective is to assist biologists in unraveling the important roles of miRNAs in human cancer; thus, it will help clinicians make sound decisions when treating cancer patients. We aim to synthesize data from existing miRNA target databases into a comprehensive conceptual model that permits an emphasis on data semantics rather than on the forms in which the data were originally represented. Consequently, a more accurate, complete view of miRNAs' biological functions can be acquired. We designed the OMIT ontology specifically for the miRNA target domain, and then carried out the semantic annotation and data integration, based upon which a domain-specific knowledge base was created. Finally, a friendly user interface was designed to demonstrate integrated information from distributed data sources, along with newly obtained knowledge via reasoning mechanisms. The overall structure of the OMIT framework is described in this section, and more details can be found in the Supplementary Material.

Overview of the OMIT Framework

As shown in Fig. 1, the main components of the OMIT framework are an ontology and a knowledge base. Information from distributed databases can be synthesized and presented to end users in a uniform view, integrated with additional information from the Gene Ontology. The Gene Ontology consists of three components (biological processes, cellular components, and molecular functions), and it provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, as well as tools to access and process such data. More details are included in the Supplementary Material.

A typical knowledge acquisition process takes eight steps:

- Steps 1 and 2: User sends a search/query to the OMIT system through the user interface
- Step 3: The recognized miRNA concept in the OMIT is used to query the knowledge base
- Step 4: miRNA targets (i.e., genes) are retrieved
- Step 5: Obtained targets are utilized to acquire more gene information
- Step 6: Related gene information is returned
- Steps 7 and 8: miRNA targets and their related gene information are returned to the user

The OMIT Ontology

The first-version OMIT ontology consists of a total of 327 concepts and 58 relationships (i.e., 28 object properties and 30 data type properties). This version has been submitted and accepted by the NCBO BioPortal. The OMIT ontology file can be freely downloaded from http://bioportal.bioontology.org/ontologies/42873

The OMIT Knowledge Base

The first-version OMIT knowledge base contains a total of 1,889 facts (referred to as "axioms" in Protégé). These facts are specified in OWL and include 27 subclass axioms, 59 disjoint class axioms, 4 sub object property axioms, 3 inverse object property axioms, 22 object property domain axioms, 27 object property range axioms, 21 data property domain axioms, 30 data property range axioms, 166 class assertion axioms, 308 object property assertion axioms, 674 data property assertion axioms, and 248 entity annotation axioms.

User Query/Search Answering

A friendly graphical user interface (GUI) to answer users' query/search has been designed with the C# language in Visual Studio 2010. As demonstrated in Fig. 2, users can specify the miRNA of interest along with expected properties of this miRNA. Both selections are made through drop-down lists so that the effort required for providing such input is minimized; corresponding values for selected properties are then retrieved and populated in a separate panel. Figure 2 exhibits part of results when "*mir-21*" and seven properties were chosen. Note that the retrieved results are regarded as integrated information in the sense that no one data source alone in our framework contains such complete knowledge. In addition to this integrated information, deep, hidden knowledge is acquired as well. Some examples include "p53 must not be a direct target of mir-885-5p" and "*mir-21* upRegulates MalignantNeoplasm." The ability to obtain previously implicit knowledge is due to the inference mechanisms applied to the knowledge base. More detailed discussion on obtaining hidden, critical domain knowledge can be found in the Supplementary Material.

CONCLUSION

In this paper, we propose an innovative computing framework based on the miRNA-domain-specific knowl-

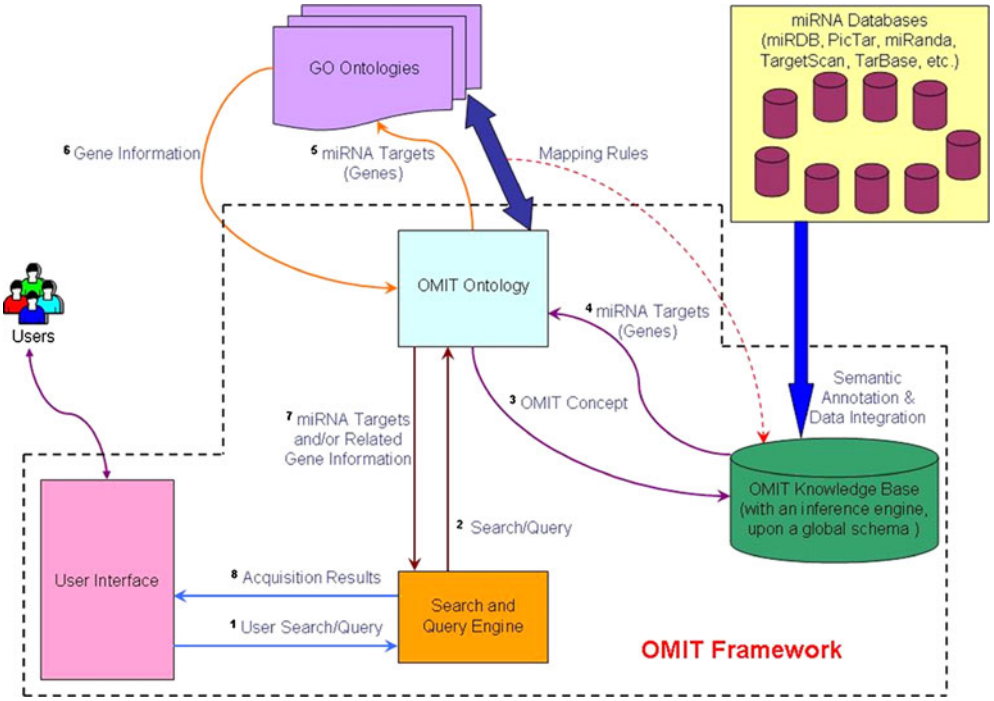

Fig. 1 Overall structure of the OMIT framework.

edge base, OMIT, to handle the challenge of an efficient acquisition of miRNAs' candidate target genes. To the best of our knowledge, the OMIT framework is designed upon the very first ontology in the miRNA domain and includes a domain-specific knowledge base. We adopt a combination of both top-down and bottom-up approaches when designing the OMIT ontology. A deep annotation is utilized during semantic annotation and data integration, which together lead to a centralized knowledge base. The OMIT system is able to assist biologists in unraveling the important roles for miRNAs in human cancer; thus, it will help clinicians make sound decisions when treating cancer patients. This long-term research goal will be achieved via facilitating knowledge discovery and sharing from existing sources.

ACKNOWLEDGMENTS

The authors would like to thank Hardik Shah and Robert Rudnick for helping in software implementation. The authors also appreciate the discussion with Patrick Hayes, Lei He, Wen-chang Lin, Hao Sun, and Xiaowei Wang.

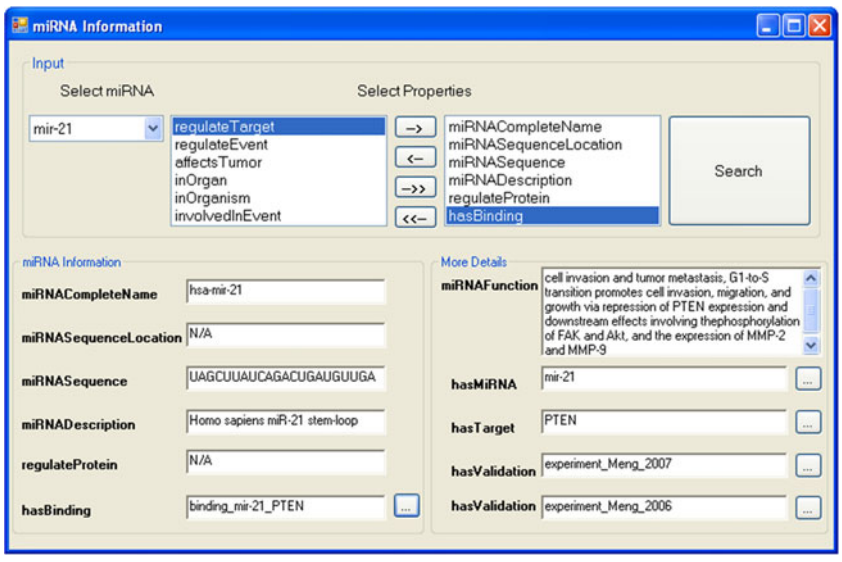

Fig. 2 Search/query GUI in the OMIT.

REFERENCES

1. Kobayashi T, Lu J, Cobb BS, Rodda SJ, McMahon AP, Schipani E, *et al.* Dicer-dependent pathways regulate chondrocyte proliferation and differentiation. Proc Natl Acad Sci. 2008; 105:1949–54.
2. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature. 2000;403:901–6.
3. Zhou M, Liu ZX, Zhao YH, Ding Y, Liu H, Xi Y, *et al.* MicroRNA-125b confers the resistance of breast cancer cells to paclitaxel through suppression of Bak1. J Biol Chem. 2010;285(28):21496–507.
4. Nakajima G, Hayashi K, Xi Y, Kudo K, Uchida K, Takasaki K, *et al.* Non-coding microRNAs hsa-let-7 g and hsa-miR-181b are associated with chemoresponse to S-1 in colon cancer. Cancer Genomics Proteomics. 2006;3:317–24.