BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 20 2013, pages 2657–2658
doi:10.1093/bioinformatics/btt439

Databases and ontologies

Advance Access publication September 6, 2013

targetHub: a programmable interface for miRNA–gene interactions

Ganiraju Manyam[1], Cristina Ivan[2], George A. Calin[3] and Kevin R. Coombes[1,*]

[1]Department of Bioinformatics and Computational Biology, [2]Department of Gynecologic Oncology and Reproductive Medicine and [3]Department of Experimental Therapeutics, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: With the expansion of high-throughput technologies, understanding different kinds of genome-level data is a common task. MicroRNA (miRNA) is increasingly profiled using high-throughput technologies (microarrays or next-generation sequencing). The downstream analysis of miRNA targets can be difficult. Although there are many databases and algorithms to predict miRNA targets, there are few tools to integrate miRNA–gene interaction data into high-throughput genomic analyses.

Results: We present targetHub, a CouchDB database of miRNA–gene interactions. TargetHub provides a programmer-friendly interface to access miRNA targets. The Web site provides RESTful access to miRNA–gene interactions with an assortment of gene and miRNA identifiers. It can be a useful tool to integrate miRNA target interaction data directly into high-throughput bioinformatics analyses.

Availability: TargetHub is available on the web at http://app1.bioinformatics.mdanderson.org/tarhub/_design/basic/index.html.

Contact: coombes.3@osu.edu

Received on March 28, 2013; revised on July 10, 2013; accepted on July 29, 2013

1 INTRODUCTION

Understanding genome-level data is a recurring task in molecular biology research, made more challenging with the development of numerous high-throughput technologies. One interesting aspect of genome biology involves microRNAs (miRNAs), which are gaining significant attention for their regulatory roles in human pathologies (Mendell and Olson, 2012). miRNAs are frequently profiled using high-throughput technologies, such as microarrays or next-generation sequencing (Li and Ruan, 2009; Ryu *et al.*, 2011). The downstream analysis of miRNA data generated from such massive platforms needs efficient bioinformatics methods.

miRNAs regulate the expression of protein-coding genes by inhibiting translation and, in some cases, by cleaving the associated messenger RNA (mRNA) (Bartel, 2004). Many computational methods predict such regulation based on sequence features, such as complementarity and free energy, of miRNA, mRNA and their duplex (Thomas *et al.*, 2010). Potential miRNA targets are often identified by seeking agreement among multiple prediction methods for a particular interaction. Currently, several databases, including TarBase (Vergoulis *et al.*, 2012), miRecords (Xiao *et al.*, 2009), miRGator (Cho *et al.*, 2011) and miRGen (Alexiou *et al.*, 2010), contain miRNA target predictions from multiple methods. The Web sites of these databases provide graphical interfaces to manually access and search the data. However, for a complete miRNA high-throughput analysis, it is currently necessary to download the complete database of target information and write customized scripts to parse the data, as there is no support for programmatic access to subsets of the miRNA–gene interaction data.

Here we present targetHub, a new programmer-friendly database of miRNA–gene interactions. TargetHub is a schema-free database that can be accessed as a web service from any modern programming language using a simple HyperText Transfer Protocol (HTTP) call. This interface supports automatic integration of miRNA target information into high-throughput miRNA data analyses. The targetHub Web site can also be used to manually access and search the miRNA–gene interactions with miRNA identifiers and an assortment of gene identifiers powered by geneSmash (Manyam *et al.*, 2012).

2 APPLICATION INFORMATION

2.1 Data description

TargetHub currently contains miRNA–gene interactions predicted by TargetScan (Grimson *et al.*, 2007), PicTar (Krek *et al.*, 2005) and miRanda (Enright *et al.*, 2003). TargetHub also includes experimentally validated interactions from miRTarBase (Hsu *et al.*, 2011). Human miRNA data were obtained from miRBase (version 18) (Griffiths-Jones, 2004). 3′ Un-Translated Region (UTR) sequence data of the human genome (hg19) were extracted from the UCSC Table Brower (Karolchik *et al.*, 2004).

Interactions predicted by TargetScan (v6.1) were obtained from their portal (www.targetscan.org). PicTar interaction data were acquired from the UCSC Table Browser (hg17). The source code of miRanda (version 3.3a) was obtained from their Web site (cbio.mskcc.org/microrna_data/manual.html) and used to compute miRNA targets for version 18 of miRBase using strict mode and the default cutoff score (140). The human miRTarBase data (r2.5) were downloaded from their portal (mirtarbase.mbc.nctu.edu.tw). Because the nomenclature for miRNA is not completely standard across these sources, miRNA names are represented in targetHub using the following conventions. miRNA that match version 18 of miRBase are retained unchanged; candidates that have no matching identifier are curated by mapping through identifiers of previous miRBase versions.

*To whom correspondence should be addressed.

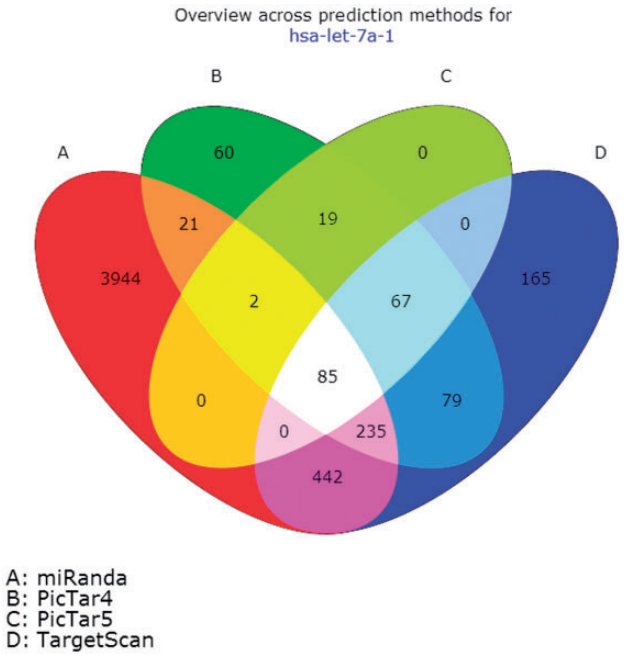

Fig. 1. Overview of the number of target genes across prediction methods for human miRNA hsa-let-7a-1 illustrated using a Venn diagram by targetHub

2.2 Implementation details

TargetHub is built using CouchDB, an Apache open-source database platform. CouchDB is a document-oriented and schema-free database system with a built-in web server. Database queries are processed through HTTP requests, which are handled by a RESTful JavaScript Object Notation (JSON) application programming interface (API) (Lennon, 2010). CouchDB provides native support for incremental database replication. Thus, users of targetHub can easily maintain a local copy with automatic updates.

The standard data document in targetHub represents an interaction between one miRNA and one gene, along with the evidence used to predict the interaction. Because query responses are pre-computed for each document, the data access rate is fast. miRNA–gene interactions predicted by other methods can easily be incorporated into targetHub, as the design/schema is flexible. The database design and source code can be freely accessed from the design document of the targetHub database.

The Web site is developed using HyperText Markup Language (HTML) and JavaScript to build on the CouchDB JSON API. The search interface relies on geneSmash, a gene-centric CouchDB database (Manyam *et al.*, 2012). The geneSmash web service converts any input gene identifier into the corresponding Entrez Gene identifier. The Entrez Gene identifier is forwarded to targetHub to retrieve the relevant miRNA–gene interaction data. Targets predicted by various methods are illustrated by a Venn diagram after searching the Web site using any criteria (Fig. 1).

The database can be accessed as a web service with an API or through the Web site. The API is described on the Web site. Sample code to access targetHub through perl and R is provided in the documentation. User can query the database with various identifiers (by either gene or miRNA) in two different search modes. Search can be performed either with aggregate number of methods supporting the interaction or by specific methods used to derive the miRNA–gene interaction.

3 CONCLUSION

TargetHub offers a Web site and web service to access miRNA–gene interactions. It is a RESTful database with a programmer-friendly interface through the API. Maintaining a local copy of targetHub is hassle-free, due to the built-in support for replication. In summary, targetHub can be a useful tool to integrate miRNA target information into high-throughput genomic analyses.

Funding: The National Cancer Institute of the National Institutes of Health (P30 CA016672 and P50 CA070907) (in part).

Conflict of Interest: none declared.

REFERENCES

Alexiou,P. *et al.* (2010) miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Res.*, **38**, D137–D141.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

Cho,S. *et al.* (2011) miRGator v2.0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Res.*, **39**, D158–D162.

Enright,A.J. *et al.* (2003) MicroRNA targets in *Drosophila*. *Genome Biol.*, **5**, R1.

Griffiths-Jones,S. (2004) The microRNA registry. *Nucleic Acids Res.*, **32**, D109–D111.

Grimson,A. *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.

Hsu,S.D. *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.

Karolchik,D. *et al.* (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.

Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

Lennon,J. (2010) *Beginning CouchDB*. Apress, New York, NY, USA.

Li,W. and Ruan,K. (2009) MicroRNA detection by microarray. *Anal. Bioanal. Chem.*, **394**, 1117–1124.

Manyam,G. *et al.* (2012) Relax with CouchDB - into the non-relational DBMS era of bioinformatics. *Genomics*, **100**, 1–7.

Mendell,J.T. and Olson,E.N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, **148**, 1172–1187.

Ryu,S. *et al.* (2011) Discovery of novel human breast cancer microRNAs from deep sequencing data by analysis of pri-microRNA secondary structures. *PLoS One*, **6**, e16403.

Thomas,M. *et al.* (2010) Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.*, **17**, 1169–1174.

Vergoulis,T. *et al.* (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.*, **40**, D222–D229.

Xiao,F. *et al.* (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.