BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 15 2014, pages 2237–2238
doi:10.1093/bioinformatics/btu155

Databases and ontologies

Advance Access publication March 20, 2014

OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs

Dongfang Wang, Jin Gu*, Ting Wang and Zijian Ding

MOE Key Laboratory of Bioinformatics, TNLIST Bioinformatics Division / Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: MicroRNAs (miRNAs), a class of small regulatory RNAs, play important roles in cancer initiation, progression and therapy. MiRNAs are found to regulate diverse cancer-related processes by targeting a large set of oncogenic and tumor-suppressive genes. To establish a high-confidence reference resource for studying the miRNA-regulated target genes and cellular processes in cancer, we manually curated 2259 entries of cancer-related miRNA regulations with direct experimental evidence from ~9000 abstracts, covering more than 300 miRNAs and 829 target genes across 25 cancer tissues. A web-based portal named oncomiRDB, which provides both graphical and text-based interfaces, was developed for easily browsing and searching all the annotations. It should be a useful resource for both the computational analysis and experimental study on miRNA regulatory networks and functions in cancer.

Availability and implementation: http://bioinfo.au.tsinghua.edu.cn/oncomirdb/

Contact: jgu@tsinghua.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 1, 2013; revised on March 14, 2014; accepted on March 17, 2014

1 INTRODUCTION

MicroRNAs (miRNAs) are a class of ~22-nt endogenous small regulatory RNAs, which can repress the expressions and/or translations of hundreds of target genes by binding to the 3′-UTRs of target gene mRNA transcripts (Bartel, 2004, 2009).Till now, there are more than 2500 human miRNAs collected in the latest miRBase database release (Griffiths-Jones, 2004; Kozomara and Griffiths-Jones, 2011). More than half of the protein-coding genes are potentially regulated by one or multiple miRNAs according to the genome-wide computational predictions and high-throughput experimental screens (Chi *et al.*, 2009; Krek *et al.*, 2005; Lewis *et al.*, 2005). Extensive evidence shows that miRNAs play important roles in nearly all essential cellular processes, such as cell proliferation, migration, differentiation, apoptosis and senescence.

Cancer, which is one of the leading causes of death worldwide, involves significant changes of chromatin structures and molecular regulatory networks. As their impacts on mRNA stability and translation efficiency are extensive, miRNAs are found to regulate different hallmarks of cancer by targeting a large set of oncogenic and tumor-suppressive genes. These miRNAs are usually called onco-miRNAs (oncomiRs) (Esquela-Kerscher and Slack, 2006; Lujambio and Lowe, 2012).

To facilitate the study on miRNA functions and regulatory networks in cancer, there is a great need to establish a reference database for annotating the oncomiRs regulating different cellular processes and target genes in different types of cancers. There are a few existing databases aiming at annotating the cancer-related miRNAs, such as miRCancer (Xie *et al.*, 2013), PhenomiR (Ruepp *et al.*, 2010), miR2Disease (Jiang *et al.*, 2009), HMDD (Lu *et al.*, 2008) and somamiR (Bhattacharya *et al.*, 2013), but they mainly collect the differentially expressed miRNAs or the miRNA-associated genetic mutations in cancer without direct functional evidence (miR2Disease and HMDD include some experimentally verified entries, but the number of entries is relatively small).

According to the PubMed records as of March 2013, there are ~9000 publications studying the cancer-related miRNAs. This literature provides extensive information for the oncomiR functions and target genes. We manually reviewed the abstracts and curated 2259 entries of experimentally verified oncomiRs, which either regulate one or more cancer-related cellular processes or directly target at least one gene in cancer-related processes. The miRNA expression patterns, their upstream regulators and the corresponding experimental conditions are also collected from the abstracts. Then, we developed a Web-based portal oncomiRDB for storing and displaying all the curated data entries. With the graphical and text-based interfaces, users can easily browse and search all the entries for the oncomiRs related to different cancer types, different regulated cellular processes or different target genes. OncomiRDB should be a valuable resource for both the computational analysis and the experimental study of miRNA functions and regulatory networks in cancer.

2 DATABASE CONTENT AND STATISTICS

To annotate the experimentally verified oncomiRs, firstly we retrieved the abstracts studying the miRNA regulations in cancer from PubMed (about 9000 abstracts for the current database release; the query details are provided in the Supplementary Material). We carefully reviewed all the abstracts and curated the detailed information of the oncomiRs with direct experimental evidence: the miRNAs regulate at least one cancer-related cellular process confirmed by perturbing their activities, such as overexpression using miRNA mimics or repression using antagomirs

*To whom correspondence should be addressed.

in cancer cell lines, and/or the miRNAs directly target at least one gene in cancer-related processes validated by luciferase reporter assay (when the validation method was not described clearly, we further checked it within the full text). In the database content, besides the information of miRNA-regulated cellular processes and target genes, the miRNA expression patterns, the miRNA upstream regulators, the cancer types and the corresponding experimental conditions were also collected if they provided in the abstracts. Because of the complexity of cancer types, the detailed cancer types were grouped into 25 different cancer tissues: for example, 'non-small cell lung cancer', 'small cell lung cancer' and 'lung adenocarcinoma' were grouped into 'lung' tissue. The latest database release contains 2259 entries covering more than 300 miRNAs and 829 target genes in 25 cancer tissues. To maintain the data quality, each entry was double-checked by at least two different curators.

For the latest release of oncomiRDB, breast, liver, colorectum, lung and stomach are the most annotated cancer tissues; miR-21, let-7a, miR-34a and miR-145 are the most common oncomiRs annotated across more than 15 cancer tissues. A few others are only studied in specific cancer tissues, such as miR-122 in liver cancer (specifically expressed in liver tissues) (Chang *et al*., 2004; Jopling *et al*., 2005). The cancer tissues and their related miRNAs form a complex network. A set of oncogenes such as CCND1, IGF1R, CDK6, MET and BCL2 is strongly regulated by multiple tumor-suppressive miRNAs. These oncogenes are frequently overexpressed in cancer tissues, consistent with the global reduction of miRNA expressions (Lu *et al*., 2005). Several chromatin modifiers, such as EZH2 and BMI1, are also extensively regulated by several miRNAs, which suggests that miRNAs play important roles in regulating the chromatin structure changes during the cancer initiation and progression.

3 USER INTERFACES

To facilitate the database browsing and searching, we built a Web-based portal providing both graphical and text-based interfaces. The graphical interface for the 'cancer tissue–miRNA' network is provided on the database main page, while the 'miRNA–target' network can be generated according to any search result. All the graphical interfaces are developed based on the Cytoscape Web utility (Lopes *et al*., 2010).

For the text-based interface, users can specify one or multiple query conditions in the left navigation column: (i) 'MicroRNA': the users can input the exact miRNA name (such as miR-21) or add '%' for a fuzzy search (such as miR-200% for miR-200a, miR-200b and miR-200c); (ii) 'MiRNA_ID': the users can select any interested miRNA with an official symbol; (iii) 'Tissue': specify one of the 25 cancer tissues; (iv) 'Tumor': specify any interested cancer subtype; (v) 'Target': specify any interest target gene; (vi) 'Function': specify one of the six most annotated cancer-related cellular functions. After submitting the query, all the matched entries will be listed. Then, the users can either generate the miRNA–target network or download the detailed annotations for listed entries. A more detailed user guide can be found in the Supplementary Material.

4 DISCUSSION

OncomiRDB is a unique resource for annotating the experimentally verified cancer-related miRNAs with direct functional evidence. By comparing with several existing miRNA target databases, oncomiRDB can provide many new miRNA–target annotations validated by luciferase assay, although it focuses on collecting cancer-related miRNA targets (the comparison result is presented in Supplementary Table S1, and the detailed list can be downloaded from the database Web site). Except for the direct targets, oncomiRB also collects the corresponding miRNA cellular functions and experimental conditions from the literature. The graphical interface, a unique feature of oncomiRDB, can achieve better database browsing and search result visualization. In summary, oncomiRDB is a high-confidence reference resource for studying the miRNA-regulated target genes and cellular processes in cancer.

Funding: National Basic Research Program of China [2012CB316503]; National Natural Science Foundation of China [61005040, 61370035, 61105003]; Tsinghua National Laboratory for Information Science and Technology Cross-discipline Foundation.

Conflict of Interest: none declared.

REFERENCES

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

Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

Bhattacharya,A. *et al*. (2013) SomamiR: a database for somatic mutations impacting microRNA function in cancer. *Nucleic Acids Res*., **41**, D977–D982.

Chang,J. *et al*. (2004) miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol*., **1**, 106–113.

Chi,S.W. *et al*. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.

Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.

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

Jiang,Q. *et al*. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*., **37**, D98–D104.

Jopling,C.L. *et al*. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*, **309**, 1577–1581.

Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*., **39**, D152–D157.

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

Lewis,B.P. *et al*. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Lopes,C.T. *et al*. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

Lu,J. *et al*. (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.

Lu,M. *et al*. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.

Lujambio,A. and Lowe,S.W. (2012) The microcosmos of cancer. *Nature*, **482**, 347–355.

Ruepp,A. *et al*. (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol*., **11**, R6.

Xie,B. *et al*. (2013) miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*, **29**, 638–644.