



Database tool

LncCeRBase: a database of experimentally validated human competing endogenous long non-coding RNAs

Cong Pian^{1,†}, Guangle Zhang^{2,†}, Tengfei Tu³, Xiangyu Ma⁴ and Fei Li^{1,*}

¹Ministry of Agriculture Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China, ²Department of Mathematics, College of Science, Nanjing Agricultural University, Nanjing, Jiangsu, China, ³Network Security Research Group, Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China and ⁴School of Life Science, Anhui Agriculture University, Hefei 230036, China

*Corresponding author: Tel.: +86 182 688 17616; Fax: +86 057 188 982 679; Email: lifei18@zju.edu.cn

[†]These authors contributed equally to this work.

Citation details: Pian, C., Zhang, G., Tu, T. *et al.* LncCeRBase: a database of experimentally validated human competing endogenous long non-coding RNAs. *Database* (2018) Vol. 2018: article ID bay061; doi:10.1093/database/bay061

Received 14 December 2017; Revised 12 April 2018; Accepted 29 May 2018

Abstract

Long non-coding RNAs (lncRNAs) are endogenous molecules longer than 200 nucleotides, and lack coding potential. lncRNAs that interact with microRNAs (miRNAs) are known as a competing endogenous RNAs (ceRNAs) and have the ability to regulate the expression of target genes. The ceRNAs play an important role in the initiation and progression of various cancers. However, until now, there is no a database including a collection of experimentally verified, human ceRNAs. We developed the LncCeRBase database, which encompasses 432 lncRNA–miRNA–mRNA interactions, including 130 lncRNAs, 214 miRNAs and 245 genes from 300 publications. In addition, we compiled the signaling pathways associated with the included lncRNA–miRNA–mRNA interactions as a tool to explore their functions. LncCeRBase is useful for understanding the regulatory mechanisms of lncRNA.

Database URL: <http://www.insect-genome.com/LncCeRBase>

Introduction

The majority of sequences in the human transcriptome are classified as lncRNA (long non-coding RNA). When compared with genes encoding proteins and small molecule RNAs (such as miRNA), the number of lncRNAs is greatest (1–3), and their regulatory mechanisms are more

diverse and extensive (4). Pre-existing evidence has shown that lncRNAs can regulate the expression of genes by interacting with proteins, RNA and DNA (5).

lncRNAs are directly involved in the regulation of gene expression and can affect an abundant number of target genes by interacting with sponging miRNAs (6). Although

the structures of most lncRNAs and mRNAs are very similar, the regulation patterns of gene expression are more diverse and wide for lncRNAs. Increasing evidence has indicated that lncRNAs play critical roles in the biological processes of cancers (7, 8). Additionally, many studies have shown that competing lncRNAs play an important role in the initiation and progression of many cancers (9–11). lncRNAs have important potential applications, including prospects for new diagnostic methods and the treatment of malignant tumors (12).

Some lncRNA databases have been constructed, including lncRNAdb (13), lncRNAWiki (14), NONCODE (15) and LNCipedia (16). Additionally, there is the DIANA-lncBase (17), which integrates miRNA–lncRNA associations. Furthermore, the databases, lncRNADisease (18), lncRNASNP (19) and lincSNP (20) are a collection of relationships between lncRNAs and diseases. These databases are crucial for exploring the functions of lncRNAs in complex diseases in humans.

Only a limited number of lncRNAs have been validated by molecular experimentation. The experimentally verified lncRNAs are highly reliable and are important references for understanding the functions of lncRNAs. However, there is no database that is devoted to collecting experimentally verified competing endogenous RNAs (ceRNAs) (lncRNA–miRNA–mRNA). Here, we developed a database (LncCeRBase) to collect experimentally supported lncRNA–miRNA–mRNA interactions. All of the triplet interactions in the LncCeRBase were manually curated from published literature. The LncCeRBase database contains 432 lncRNA–miRNA–miRNA interactions, including 130 lncRNAs, 214 miRNAs and 245 genes from 300 publications. The LncCeRBase database should be helpful in understanding the regulatory mechanisms of lncRNA in complex diseases.

Data sources and implementation

In constructing the database, scientific publications were obtained through searching keywords such as ‘lncRNA’, ‘ceRNA’, ‘competing RNA’, ‘lncRNAs targeting’ ‘lncRNA targeting’, ‘miRNA sponges’ and ‘circRNAs as miRNA sponges’ in the PubMed database of the National Center for Biotechnology Information (NCBI). Then, we selected the resulting literature describing lncRNA–miRNA–mRNA triplet interactions. All of the selected lncRNA–miRNA–mRNA interactions were experimentally confirmed by utilizing RNAi, Western blots, qRT-PCR or luciferase reporter assays.

In addition, many lncRNAs can activate or inhibit signaling pathways as ceRNAs by sponging miRNAs (21, 22). Therefore, we compiled the signaling pathways of the

mRNAs involved in lncRNA–miRNA–mRNA associations. Finally, the LncCeRBase database contains 432 lncRNA–miRNA–miRNA interactions, including 130 lncRNAs, 214 miRNAs and 245 genes from 300 publications.

MongoDB is a free and open-source cross-platform document-oriented database program. When compared with the classic MySQL, MongoDB has the following five advantages: (i) Weak consistency; (ii) The way the document structure is stored, the data can be accessed more easily; (iii) The built-in GridFS supports larger capacity storage; (iv) Built-in Sharding; and (v) Third parties are rich in support (this is the advantage of MongoDB compared with other NoSQL) (23). Therefore, all data in the LncCeRBase were stored and managed using MongoDB (version 3.2.). The web interfaces were built-in Python (version 3.5). The data processing programs were written in Python (version 3.5), and the web services were built using Nginx.

Web interface

The web service, LncCeRBase, is available at lncCeRBase.insect-genome.com. Users can browse the lncRNA names, miRNA names, mRNA gene names or diseases. When selecting an lncRNA, miRNA or mRNA in the ‘Browse’ page of the web site, the LncCeRBase will return a list of matched lncRNA–miRNA–mRNA triplet associations, containing the name (lncRNA, miRNA, and mRNA), PubMed ID, associated disease/tissue, description, title and pathway name. For every entity, we link the name of lncRNA, miRNA and mRNA to the resource of RNAcentral (24), miRBase (25) and NCBI gene (26) respectively. Besides, all data from the LncCeRBase database can be downloaded. Since a gene may have other names, we design the search section; users can determine lncRNA–miRNA–mRNA triplet associations by inputting any name of a gene.

The application of LncCeRBase

LncCeRBase provides a user-friendly interface to conveniently browse, search and download data. With the rapidly increasing interest in ceRNA, LncCeRBase will significantly improve our understanding of lncRNA–miRNA–mRNA triplet associations in diseases and has the potential to be a valuable resource.

Future directions

The LncCeRBase database will be updated with new experimentally supported lncRNA–miRNA–miRNA interactions every two months. We found that 90.5% (391/432) of the included lncRNA–miRNA–miRNA interactions

were verified between 2016 and 2017. This phenomenon shows that the regulatory mechanisms of competing endogenous lncRNAs have recently been gaining increasing attention. Undoubtedly, there will be many studies regarding competing endogenous lncRNAs in the future. In recent years, several ceRNA prediction methods were proposed. For example, Sardina *et al.* developed a computational method, called CERNIA, which takes into account insights from *in vivo* and *in silico* experiments, such as 5' UTR and coding region binding sites, and tissue-specific gene expression profiles, to uncover novel ceRNAs, by taking into account both validated and high-confidence miRNA–target interactions (27). Zarringhalam *et al.* predicted the ceRNA network of PTEN by calculating a set of probabilistic features (28). Zhang *et al.* proposed a multi-step method called miRSCoPPI to infer miRNA sponge co-regulation of protein–protein interactions in the breast cancer (29). However, the studies of ceRNA prediction algorithm are limited. We are also developing an algorithm to predict lncRNA–miRNA–miRNA interactions by constructing an lncRNA–miRNA–mRNA network. With the generation of large-scale RNA-Seq data from TCGA (The Cancer Genome Atlas), increasing lncRNA–miRNA–miRNA interactions will be discovered. This tool will be based on mRNA, lncRNA and miRNA expression data found in TCGA and will be integrated into the LncCeRBase in the near future.

Discussion and conclusion

Here, we developed a database (LncCeRBase) to collect experimentally supported lncRNA–miRNA–mRNA interactions. The LncCeRBase database integrates this triplet interaction data and can help us to explore the regulatory mechanisms of lncRNAs. With the increasing attention and deepening of research on lncRNA genes, an increasing number of new lncRNAs have been discovered. Currently, the functions of many lncRNAs are unknown, and the function of lncRNAs as ceRNAs is a research area that has been even less explored. Therefore, the lncRNAs, which have a biological function as ceRNAs, deserve investigation.

Funding

This work was supported by the National Key Research and Development Program (2016YFC1200600, 2017YFC1200602).

Conflict of interest. None declared.

References

- Bo, H., Gong, Z., Zhang, W. *et al.* (2015) Upregulated long non-coding RNA AFAP1-AS1 expression is associated with progression and poor prognosis of nasopharyngeal carcinoma. *Oncotarget*, **6**, 20404–20418.
- Zhang, W., Fan, S., Zou, G. *et al.* (2015) Lactotransferrin could be a novel independent molecular prognosticator of nasopharyngeal carcinoma. *Tumour Biol.*, **36**, 675–683.
- Yan, Q., Zeng, Z., Gong, Z. *et al.* (2015) EBV-miR-BART10-3p facilitates epithelial-mesenchymal transition and promotes metastasis of nasopharyngeal carcinoma by targeting BTRC. *Oncotarget*, **6**, 41766–41782.
- Geisler, S., and Collier, J. (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.*, **14**, 699–712.
- Wang, K.C., and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904–914.
- Salmena, L., Poliseno, L., Tay, Y. *et al.* (2011) A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell*, **146**, 353–358.
- Yuan, J.-h., Yang, F., Wang, F. *et al.* (2014) A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*, **25**, 666–681.
- Prensner, J.R. *et al.* (2013) The long noncoding RNA SchLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.*, **45**, 1392–1398.
- Wang, J., Liu, X., Wu, H. *et al.* (2010) CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.*, **38**, 5366–5383.
- Liu, X.-H., Sun, M., Nie, F.-Q. *et al.* (2014) LncRNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol. Cancer*, **13**, 92–117.
- Zhou, X., Gao, Q., Wang, J. *et al.* (2014) Linc-RNA-RoR acts as a “sponge” against mediation of the differentiation of endometrial cancer stem cells by microRNA-145. *Gynecol. Oncol.*, **133**, 333–339.
- Zhang, W., Huang, C., Gong, Z. *et al.* (2013) Expression of LINC00312, a long intergenic non-coding RNA, is negatively correlated with tumor size but positively correlated with lymph node metastasis in nasopharyngeal carcinoma. *J. Mol. Histol.*, **44**, 545–554.
- Quek, X.C., Thomson, D.W., Maag, J.L.V. *et al.* (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, **43**, D168–D173.
- Ma, L., Li, A., Zou, D. *et al.* (2015) LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.*, **43**, D187–D192.
- Xie, C., Yuan, J., Li, H. *et al.* (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.*, **42**, D98–D103.
- Volders, P.-J., Verheggen, K., Menschaert, G. *et al.* (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.*, **43**, D174–D180.
- Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N. *et al.* (2013) DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.*, **41**, D239–D245.
- Chen, G., Wang, Z., Wang, D. *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983–D986.

19. Gong,J., Liu,W., Zhang,J. *et al.* (2015) lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.*, **43**, D181–D186.
20. Ning,S., Zhao,Z., Ye,J. *et al.* (2014) LincSNP: a database of linking disease-associated SNPs to human large intergenic non-coding RNAs. *BMC Bioinformatics*, **15**, 152.
21. Poliseno,L., Salmena,L., Zhang,J. *et al.* (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.
22. Xia,T., Liao,Q., Jiang,X. *et al.* (2014) Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. *Sci. Rep.*, **4**, 6088.
23. Kumar1,L. *et al.* (2015) Comparative analysis of NoSQL (MongoDB) with MySQL Database. *Sci. J. Impact Factor*, **5**, 2349–9745.
24. The RNAsentral consortium. (2017) RNACentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.*, **45**, D128–D129.
25. Kozomara,A., and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
26. Brown,G.R., Hem,V., Katz,K.S. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
27. Sardina,D.S. *et al.* (2017) A novel computational method for inferring competing endogenous interactions. *Brief. Bioinform.*, **6**, 1071–1081.
28. Zarringhalam,K., Tay,Y., Kulkarni,P. *et al.* (2017) Identification of competing endogenous RNAs of the tumor suppressor gene PTEN: a probabilistic approach. *Sci. Rep.*, **7**, 7755.
29. Zhang,J.P. *et al.* (2017) Inferring miRNA sponge co-regulation of protein-protein interactions in human breast cancer. *BMC Bioinformatics*, **18**, 243.