BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 17 2013, pages 2221–2222
doi:10.1093/bioinformatics/btt361

Databases and ontologies

Advance Access publication June 22, 2013

Linc2GO: a human LincRNA function annotation resource based on ceRNA hypothesis

Ke Liu†, Zhangming Yan†, Yuchao Li and Zhirong Sun*

MOE Key Laboratory of Bioinformatics, State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Life Sciences, Tsinghua University, Beijing 100084, China

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: Large numbers of long intergenic non-coding RNA (lincRNA) have been detected through high-throughput sequencing technology. However, currently we still know very little about their functions. Therefore, a lincRNA function annotation database is needed to facilitate the study in this field. In this article, we present Linc2GO, a web resource that aims to provide comprehensive functional annotations for human lincRNA. MicroRNA-mRNA and microRNA-lincRNA interaction data were integrated to generate lincRNA functional annotations based on the 'competing endogenous RNA hypothesis'. To the best of our knowledge, Linc2GO is the first database that makes use of the 'competing endogenous RNA hypothesis' to predict lincRNA functions.

Availability: Freely available at http://www.bioinfo.tsinghua.edu.cn/~liuke/Linc2GO/index.html

Contact: sunzhr@mail.tsinghua.edu.cn

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

Received on October 24, 2012; revised on May 31, 2013; accepted on June 18, 2013

1 INTRODUCTION

Long intergenic non-coding RNAs (lincRNAs) are endogenous long non-coding RNA molecules that transcribed from 'intergenic' regions of the genome. In recent years, as high-throughput sequencing technology developed, more and more lincRNA have been identified (Guttman *et al.*, 2009; Pauli *et al.*, 2012; Young *et al.*, 2012).

It has been demonstrated that lincRNAs play critical roles in regulating multiple important biological processes (Guttman *et al.*, 2011; Guttman and Rinn, 2012; Khalil *et al.*, 2009; Moran *et al.*, 2012). However, currently most lincRNAs are still poorly studied. Owing to the importance of lincRNA, it has become very necessary to develop computational tools to predict lincRNA functions. Previous researchers have proposed a 'co-expression-based' method to predict lincRNA functions and achieved many meaningful results (Guo *et al.*, 2013; Liao *et al.*, 2011; Loewer *et al.*, 2010). That is, if a lincRNA is co-expressed with a protein-coding gene whose function is already known, then the lincRNA is predicted to take similar functions.

In this article, we predict lincRNA functions in a rather different approach and present Linc2GO, a novel lincRNA function annotation database. Instead of using the co-expression information of lincRNA and protein coding-gene, our work is based on the competing endogenous RNA hypothesis (ceRNA hypothesis): lincRNA can function as microRNA 'sponge' to interact directly with microRNA and prevent them from binding to mRNA. In this way, lincRNA regulates gene expression and meanwhile regulates the biological processes in which they get involved in (Salmena *et al.* 2011; Zhao *et al.*, 2008). For example, MAML1 and MEF2C are two important transcription factors that activate muscle-specific gene expression. Marcella Cesana *et al.* showed that a lincRNA, linc-MD1, regulates muscle differentiation by interacting with two microRNAs, miR-135 and miR-133, which can bind to MAML1 and MEF2C to regulate their expressions (Cesana *et al.*, 2011). Based on the above fact, it is natural to infer that if a lincRNA and an mRNA share some microRNAs that can interact with both of them, then the two have similar biological functions.

2 METHODS

We downloaded three microRNA-mRNA interaction datasets predicted by three different algorithms: TargetScan (Bartel, 2007; Lewis *et al.*, 2005), miRanda (Betel *et al.*, 2010) and PITA (Kertesz *et al.*, 2007). Then we integrated them into one dataset, which is much more accurate (See Supplementary Material). Finally, we got 1 218 961 microRNA-mRNA interactions.

Human lincRNAs are from 'Human lincRNA Catalog' (Cabili *et al.*, 2011). All lincRNA sequences were downloaded from UCSC genome browser. MicroRNA sequences were downloaded from miRBase (Kozomara and Griffiths-Jones, 2011). The miRanda software is used to predict microRNA-lincRNA interactions with default parameters. We finally got 2 198 132 human microRNA-lincRNA interactions.

Having acquired the microRNA-mRNA and microRNA-lincRNA interaction data, we followed the principle and workflow shown in Figure 1 to generate lincRNA functional annotations.

First, for each ceRNA-ceRNA pair (the ceRNA-ceRNA pairs here include lincRNA-mRNA pairs, mRNA-mRNA pairs and lincRNA-lincRNA pairs), hypergeometric distribution was used to measure whether the two ceRNAs significantly share some microRNAs that can interact with both of them. The *P*-value was calculated as:

$$P = 1 - \sum_{t=0}^{x-1} \frac{\binom{L}{t}\binom{N-L}{M-t}}{\binom{N}{M}}$$

where N is the total number of microRNA, M is the number of microRNAs that interact with the first ceRNA, L is the number of microRNAs that interact with the second ceRNA and x is the number

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

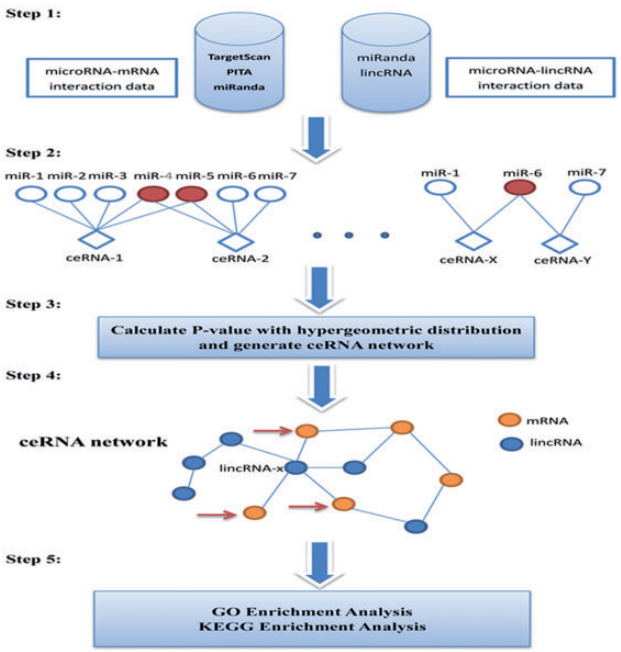

Fig. 1. Principle and workflow of generating lincRNA functional annotations. The mRNA neighbors of lincRNA-x, which would be used to predict functions of lincRNA-x in step 5, were pointed out by red arrows.

of microRNAs that interact with both of them. Only the ceRNA-ceRNA pairs with a small *P*-value (FDR <0.05) were kept for further analyses.

Next, we combined the kept ceRNA-ceRNA pairs to generate a 'ceRNA network' with network nodes represented by ceRNA (either lincRNA or mRNA), and the two ceRNAs presented in the same ceRNA-ceRNA pair were connected by an edge. For each lincRNA in the ceRNA network, we collected all its mRNA neighbors and mapped them to corresponding genes, getting a gene set. GO enrichment analysis and KEGG enrichment analysis were then applied to the gene set. The final generated GO terms and KEGG pathway terms were assigned to the lincRNA as its functional annotation.

3 RESULTS

In the finally generated functional annotations, there were 7202 lincRNAs that were assigned with at least one GO BP term. We did a detailed analysis of GO BP terms and showed that lincRNAs play critical roles in multiple biological processes such as nervous system processes, cellular development, cell-cell communication, biological adhesion, cellular component organization and so on (see Supplementary Material). As we know, Linc2GO is the first database that provides comprehensive human lincRNA function annotations.

The Linc2GO has provided a user-friendly interface to access functional annotations (See Supplementary Fig. S1). Three additional text files in csv format that contain all predicted functional annotations are also downloadable to provide convenient access in computational projects.

Funding: This work was supported by 973 projects (No. 2009CB918801, No. 2012CB705200) of China, and National Natural Science Foundation of China (NSFC) fund (No.31171274).

Conflict of Interest: none declared.

REFERENCES

Bartel,D.P. (2007) MicroRNAs: Genomics, biogenesis, mechanism, and function (Reprinted from *Cell*, **116**, 281–297, 2004). *Cell*, **131**, 11–29.

Betel,D. *et al.* (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.

Cabili,M.N. *et al.* (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.

Cesana,M. *et al.* (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA (erratum *Cell*, **147**, 948). *Cell*, **147**, 358–369.

Guo,X. *et al.* (2013) Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.*, **41**, e35.

Guttman,M. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.

Guttman,M. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.

Guttman,M. and Rinn,J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.

Kertesz,M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.

Khalil,A.M. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA*, **106**, 11667–11672.

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

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.

Liao,Q. *et al.* (2011) ncFANs: a web server for functional annotation of long noncoding RNAs. *Nucleic Acids Res.*, **39**, W118–W124.

Loewer,S. *et al.* (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.*, **42**, 1113–1117.

Moran,V.A. *et al.* (2012) Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.*, **40**, 6391–6400.

Pauli,A. *et al.* (2012) Systematic identification of long non-coding RNAs expressed during zebrafish embryogenesis. *Genome Res.*, **22**, 577–591.

Salmena,L. *et al.* (2011) A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell*, **146**, 353–358.

Young,R.S. *et al.* (2012) Identification and properties of 1,119 candidate LincRNA loci in the Drosophila melanogaster genome. *Genome Biol. Evol.*, **4**, 427–442.

Zhao,Y. *et al.* (2008) MicroRNA regulation of messenger-like noncoding RNAs: a network of mutual microRNA control. *Trends Genet.*, **24**, 323–327.