Databases and ontologies

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; Guttmann 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 (Kertesza et al., 2007). Then we integrated them into one dataset, which is much more accurate (See Supplementary Material). Finally, we got 1218961 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 2198132 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 ceRNA significantly share some microRNAs that can interact with both of them. The P-value was calculated as:

\[ P = 1 - \sum_{i=0}^{\min(L, N-M)} \frac{\binom{L}{i} \binom{N-L}{x-i}}{\binom{N}{x}} \]

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.

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com
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**


